REPORT RESUMES

ED 014 397 24 RE 000 406
IDENTIFICATION AND MEASUREMENT OF READING SKILLS OF
HIGH-SCHOOL STUDENTS.
BY- DAVIS, FREDERICK B.
PENNSYLVANIA UNIV., PHILADELPHIA
REPORT NUMBER CRP-3023 PUB DATE 67
REPORT NUMBER BR-5-0550
CONTRACT OEC-6-10-024
EDRS PRICE MF-$0.50 HC-$3.44 84P.

DESCRIPTORS- *GRADE 12, *READING COMPREHENSION, *TEST CONSTRUCTION, *READING RESEARCH, READING SKILLS, TEST RELIABILITY, READING TESTS,

A STUDY DESCRIBED AS THE FIRST APPLICATION OF CROSS-VALIDATED UNIQUENESS ANALYSIS TECHNIQUES WAS DESIGNED TO ELIMINATE THE EFFECTS OF IMPERFECTIONS IN A PRIOR FACTOR-ANALYTIC STUDY OF READING COMPREHENSION WHICH USED TESTS ESPECIALLY CONSTRUCTED TO MEASURE MENTAL SKILLS IN READING. A UNIQUENESS ANALYSIS BASED ON LARGE SAMPLES WAS USED TO OBTAIN ESTIMATES OF THE PERCENTAGE OF NONCHANCE VARIANCE. STUDENTS IN GRADE 12 NEAR PHILADELPHIA SERVED AS SUBJECTS FOR THE PRELIMINARY AND MAIN STUDIES. EIGHT READING COMPREHENSION SKILLS WERE TESTED. IN THE PRELIMINARY STUDY, TWO PARALLEL FORMS OF THE TEST WERE ADMINISTERED TO APPROXIMATELY 400 STUDENTS. A POOL OF 24 ITEMS FOR EACH OF THE EIGHT SKILLS WAS ASSIGNED TO TWO PARALLEL FORMS OF THE TEST. APPROXIMATELY 1,000 STUDENTS TOOK BOTH FORMS OF THE TEST WITH 1 OR 2 DAYS INTERVENING. INTERCORRELATIONS AND MULTIPLE REGRESSION ANALYSES WERE USED TO ANALYZE THE DATA. IT IS CONCLUDED THAT COMPREHENSION AMONG MATURE READERS IS NOT A UNITARY MENTAL SKILL OR OPERATION. IT IS POINTED OUT THAT SYSTEMATIC LEARNING EXERCISES APPROPRIATE IN LEVEL OF DIFFICULTY FOR EACH PUPIL SHOULD BE PROVIDED. IN ADDITION, THERE IS A NEED FOR A SERIES OF SELF-TEACHING PRACTICE EXERCISES FOR DEVELOPING PROFICIENCY IN THE CONSTITUENT SKILLS OF COMPREHENSION AMONG MATURE READERS. A BIBLIOGRAPHY IS GIVEN. (BK)

ED014397

IDENTIFICATION AND MEASUREMENT OF READING SKILLS OF HIGH-SCHOOL STUDENTS

Cooperative Research Project No. 3023

FREDERICK B. DAVIS
University of Pennsylvania
Philadelphia

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR 5-0550

IDENTIFICATION AND MEASUREMENT OF READING SKILLS OF HIGH-SCHOOL STUDENTS

PA 24

OEC-6-10-024

RE 000 406

Cooperative Research Project No. 3023

FREDERICK B. DAVIS
University of Pennsylvania
Philadelphia

1967

RE 000 406

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ACKNOWLEDGMENTS

The writer is grateful to the administrators, staff members, and students in the cooperating school systems of Collingdale, Radnor, Rose Tree, and Upper Darby, Pennsylvania. Specific acknowledgments are made in the text.

As Research Associate, Charlotte Croon Davis selected and edited the reading passages and items for the tryout and final forms of the tests.

Special mention should also be made of Mr. Gerald Cimmet and Mr. Lawrence Lengel, who made arrangements for the tryout and final testing sessions, of Mrs. Beatrice Mittelman, who programed the operations on the computer, and of Miss Geraldine Higgs, who typed both the test materials and this report.

Without the generous cooperation of the Educational Testing Service and the Psychological Corporation in allowing experimental use of items originally tried out by the writer for the Cooperative Reading Comprehension Tests and the Davis Reading Tests, the study could not have been carried out with a modest allotment of time and funds.

This study was made possible by a grant from the Cooperative Research Branch, Bureau of Research, U. S. Office of Education.

Frederick B. Davis
Principal Investigator

University of Pennsylvania
Philadelphia, Pennsylvania

CONTENTS

Page

Acknowledgments ii

Table of Contents iii

List of Tables. iv

Section 1: Research in Comprehension in Reading 1

Section 2: Identification and Measurement of Reading Skills of High-School Students 18

Section 3: Conclusions 73

Bibliography. 76

LIST OF TABLES

Number	Title	Page
1	Data Pertaining to Tests of Nine Skills in Reading Comprehension Used by Davis in 1940	8
2	Factor Loadings (Correlations Between Each Test and Each Factor) Found by Davis in 1941	11
3	Variance Ratios Between Successive Pairs of Factor Variances Found by Davis in 1941	12
4	Reliability Coefficients, Means, and Standard Deviations of Five Factors Found by Davis in 1941	13
5	Sample Items Measuring Eight Skills of Comprehension in Reading	20
6	Composition of Experimental Reading Tests, Forms A and B, Administered in March 1966	24
7	Distributions of Total Scores on Forms A and B	25
8	Median Correlation Coefficients of Items With Total Scores for Skills Supposed to Be Measured and for Skills Not Supposed to Be Measured	27
9	Median Percentage of Examinees Marking Items Correctly in Forms C and D	28
10	Composition of Final Reading Tests, Forms C and D, Administered in October-November 1966	30
11	Within-Day Reliability Coefficients, Means, and Variances of Skills 1-8 in Forms C and D	33
12	Across-Day Reliability Coefficients, Means, and Variances of Skills 1-8 in Form C or Form D	36
13	Intercorrelations of Skills 1-8 in Form C (Within First Testing Session)	40
14	Equations Used in Estimating Unique Nonchance Variance	44
15	Matrixes of Intercorrelations Used For Estimating Multiple Correlations	46
16	Proportion of Variance of Each of Eight Variables Accounted for by Seven Others in Original and Cross-Validated Samples	47

LIST OF TABLES (Continued)

Number	Title	Page
17	Reliability Coefficients of Reading-Skill Scores and of Cross-Validated Weighted Composite Scores (of Seven Variables Each) Used in Estimating Unique Nonchance Variance	55
18	Proportions of Unique Variance in the Nonerror Variance	
	Skill 1	62
	Skill 2	63
	Skill 3	64
	Skill 4	65
	Skill 5	66
	Skill 6	67
	Skill 7	68
	Skill 8	69
19	Summary of Percentages of Unique Variance in the Nonerror Variance of Eight Reading Skills (N = 988)	72

Section 1

RESEARCH IN COMPREHENSION IN READING

Research in reading began during the latter part of the nineteenth century with studies of eye movements; for many years thereafter investigators were largely concerned with studying the mechanical aspects of the reading process. The importance of comprehension was recognized, but it was regarded as a natural concomitant of the mechanics of reading and little direct attention was paid to it. Horace Mann told of the fluent oral reading he observed in a classroom that he was visiting. When he asked one of the pupils to read from a newspaper that he happened to have with him, the pupil read each line all the way across the page, paying no attention to the fact that the page was divided into several separate columns.

During the first half of the twentieth century, the emphasis on oral reading gradually declined, especially with respect to mechanical "word calling." The de-emphasis, in fact, went so far that, since 1950, the value of oral reading as a means of measuring an individual's ease and accuracy of converting graphemes into phonemes has had to be called to the attention of teachers of reading, especially in the lower grades.

Despite the long-standing interest in teaching reading as a thought-getting process, there has been a surprisingly small number of experimental studies on the nature of the mental skills involved in comprehension. The first important studies of this nature were published by Thorndike in 1917.[1] In three articles he analyzed

---

[1] E. L. Thorndike, "The Psychology of Thinking in the Case of Reading," Psychological Review, XXIV (1917), 162-170.

E. L. Thorndike, "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," Journal of Educational Psychology, VIII (1917), 323-332.

E. L. Thorndike, "The Understanding of Sentences," Elementary School Journal, XVIII (1917), 98-114.

---

the errors made by elementary-school pupils in writing the answers to simple questions based on short paragraphs that were presented to them. The pupils were given unlimited time and allowed to refer to the paragraphs as often as they wished, but Thorndike found that even when the pupils understood the meaning of the individual words in a paragraph, many of them made errors in answering questions about it. The nature of these errors led him to conclude that the pupils were unable to use relational words and phases (such as "but" or "on the contrary") to fit together the separate ideas expressed or to give to the individual words or word groups the proper amount of emphasis with respect to one another. An element that becomes unjustifiably dominant for a given individual he described as "over-potent"; conversely, an element that becomes unduly weak he described as "under-potent." After noting the importance of knowing the meanings of the words in a paragraph, he stated:

> "The successful response to a question or to a paragraph's meaning implies the restraint of tendencies of many words to be over-potent and the special weighting of other tendencies. This task is quite beyond the power of weak minds and is of the same selective and coordinating nature as the more obvious forms of reasoning in mathematics or science."[2]

> "Understanding a paragraph is like solving a problem in mathematics. It consists in selecting the right elements of the situation and putting them together in the right relations, and also with the right amount of weight or influence

---

[2] E. L. Thorndike, "The Understanding of Sentences," _Elementary School Journal_, XVIII (1917), 114.

or force for each."[3]

---

[3] E. L. Thorndike, "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," Journal of Educational Psychology, VIII (1917), 329.

> "Understanding a . . . printed paragraph is then a matter of habits, corrections, mental bonds, but these have to be selected from so many others, and given weights so delicately, and used together in so elaborate an organization that 'to read' means 'to think' as truly as does 'to evaluate' or 'to invent' or 'to demonstrate' or 'to verify.'"[4]

---

[4] E. L. Thorndike, "The Understanding of Sentences," Elementary School Journal, XVIII (1917), 114.

In 1925-1926 Alderman reported, in the results of a training study, that comprehension, as measured by the Thorndike-McCall Reading Scale, was improved by drill work in vocabulary building, retention exercises, and practice in organization. The latter, which consisted of selecting the central thought of each of a number of paragraphs and arranging them logically according to the writer's purpose, proved to be most effective.[5]

---

[5] G. H. Alderman, "Improving Comprehension Ability in Silent Reading," Journal of Educational Research, XIII (1926), 11-21.

The study published by T. W. H. Irion in 1925 presented data about the relationships among various scores in reading comprehension.[6] For example, ability to

---

[6] Irion, T. W. H., Comprehension Difficulties of Ninth-Grade Students in the Study of Literature. New York: Teachers College, Columbia University, 1925.

---

answer factual questions about a passage was shown to have a low correlation (.46) with ability to determine the main point and the conclusions of a writer. Irion's data also showed fairly low correlations among test scores based on different types of reading materials and indicated that word knowledge plays an important role in comprehension.

From a study of comprehension of detailed directions, Carroll concluded in 1927 that the chief sources of errors made by pupils were in

1. sentences that involve some mathematical calculations,
2. sentences that contain conditional clauses,
3. sentences that are compact or involved,
4. sentences that present material that is not explicitly stated but is merely implied.

One of the most penetrating and insightful analyses of comprehension was published by Richards in 1929.[7] Practical Criticism was devoted to the understanding of poetry, but most of its conclusions may be applied to the comprehension of prose. No summary can do justice to this important book, which should be studied carefully along with Interpretation in Teaching.[8] For adequate comprehension, Richards concludes that a reader must:

1. Understand the literal sense meaning of the writer,
2. Recognize the writer's feeling or mood,
3. Apprehend the writer's tone; that is, his attitude toward the reader,
4. Recognize a writer's intent or purpose,

---

[7] I. A. Richards, Practical Criticism. New York: Harcourt, Brace and World, 1929.

[8] I. A. Richards, Interpretation in Teaching. New York: Harcourt, Brace and World, 1938.

5. Blend correctly the four points mentioned above.

In 1931, Betty Trier Berry summarized the skills characteristic of a good reader as follows:

> "In reading to master the general outline or the facts involved, the good reader is able to outline the selection, identifying main and subordinate topics, to relate subordinate details to such an outline, to select key sentences or determine topics of paragraphs or of a longer selection, to accompany his reading with appropriate visual imagery, to note for later consideration new or difficult terms and concepts, and to grasp the major issues and their implications.
>
> "In close reading for mastery of content, the good reader is able to understand the individual words and make reasonable inferences as to meanings of words he does not know, to accept for the moment the writer's point of view, disregarding his own prejudices and biases, to give to words and phrases the meanings and interpretations intended by the writer, to follow a train of thought through a maze of detail, to ignore whatever is irrelevant for his purpose, to select and organize data for use in answering questions or the like, to isolate the essential parts of an idea, to note restrictive modifications, to group essential ideas or elements (after these have been isolated) in meaningful relationships, and to associate the selection as a unit with what precedes and what follows it."[9]

---

[9] B. T. Berry, "Improving Freshman Reading Habits," English Journal (College Edition), XX (1931), 824-828.

From her experimental results, she concluded:

> "Apparently one can leave out of consideration the question of the student's native intelligence and cultural background and can

successfully direct training to the improvement of specific and technical reading comprehension abilities."

In another article written by Berry with Touton, the results of a detailed analysis of 20,003 errors in comprehension made by 738 college entrants indicated that the errors may be classified in the following way:

1. Inability to understand fully the question to be answered,
2. Inability to isolate the elements of an involved statement read in context,
3. Inability to associate the related elements of the context,
4. Failure to grasp or retain ideas essential to the understanding of additional concepts,
5. Failure to see the setting of the context as a whole,
6. Irrelevant answers of various types,[10]

---

[10] F. C. Touton and B. T. Berry, "Reading Comprehension at the Junior College Level," California Quarterly of Secondary Education, VI (1931), 245-251.

In 1934, Dewey reported a study of the relation between ability to obtain facts and to carry out inferential thinking in historical material. He found product-moment correlation coefficients between those two variables ranging from .38 to .65 and concluded that we should not assume that tests that measure skill in obtaining facts adequately measure understanding.[11]

---

[11] J. C. Dewey, "The Acquisition of Facts as a Measure of Reading Comprehension," Elementary School Journal, XXXV (1935), 346-348.

In two articles, Feder described in 1938 the construction and use of what he called Comprehension Maturity Tests.[12] He obtained data showing that reading for information and reading for inference are relatively independent. His was the first study that involved techniques of factor analysis; he analyzed a matrix that included tests of factual reading, inference reading, appreciation, and speed of reading that involved comprehension of simple material.

---

[12] D. D. Feder, "Comprehension Maturity Tests -- A New Technique in Mental Measurement," Journal of Educational Psychology, XXIX (1938), 597-606.

In 1940-1941, Davis conducted the first factor-analytic study of comprehension that made use of tests especially constructed to measure the mental skills in reading regarded as of greatest importance by authorities in the field.[13] As a first step, a careful survey was made of the literature to identify the comprehension skills deemed most important; the result was a list of several hundred skills, many of them overlapping. In this list, nine groups of testable skills were sorted out--clusters that seemed likely to have relatively high correlations within themselves and somewhat lower correlations with other clusters.

Five-choice multiple-choice items were constructed to measure the nine basic comprehension skills listed in Table 1. The numbers of items administered to the experimental group varied from 5 to 60 (as shown in Table 1) because they formed part of Form Q of the Cooperative Reading Comprehension Tests and the study was performed as a by-product of the development of that important series of tests.[14] The means and variances of these nine skill tests in a sample of 421 college freshmen who answered each of the items in the unlimited time provided are also shown in Table 1.

As would be expected, the reliability coefficients of the nine tests varied

---

[13] F. B. Davis, "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (1944), 185-197.

[14] F. B. Davis, *et al*. The Cooperative Reading Comprehension Tests, Lower and Higher Levels, Forms Q, R, S, T, Y, Z. Princeton, N.J.: Educational Testing Service, 1940-1949.

TABLE 1

DATA PERTAINING TO TESTS OF NINE SKILLS IN READING COMPREHENSION USED BY DAVIS IN 1940*

Skill	No. of Items	Mean	Variance
1. Recalling word meanings	60	23,77	134.70
2. Drawing inferences about the meaning of a word from content	20	12.70	10.56
3. Following the structure of a passage	9	4.20	3.01
4. Formulating the main thought of a passage	5	2.97	1.22
5. Finding answers to questions answered explicitly or merely in paraphrase in the content	22	18.10	6.05
6. Weaving together ideas in the content	42	25.67	32.17
7. Drawing inferences from the content	43	28.46	33.75
8. Identifying a writer's techniques, literary devices, tone, and mood	10	6.75	3.46
9. Recognizing a writer's purpose, intent, and point of view	27	15.19	16.54

---

*Frederick B. Davis, "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (1944), 185-197.

considerably--from .17 for Skill 4 to .90 for Skill 1. This wide variation in relative precision of measurement was a major weakness of Davis's study. A second weakness was that several items, often testing different skills, were based on the same reading passage; as a result, both the intercorrelations of the scores on the nine skills and their split-half reliability coefficients were spuriously high in amounts that varied from test to test. It was primarily to eliminate the effects of these imperfections in Davis's original study (which have also affected the findings of other studies in reading comprehension) that the present study was designed and carried out.

Davis obtained the variances and covariances of the nine skill tests and performed a principal-axis analysis of the resulting matrix. Since the diagonal entries were variances, the analysis included not only the common variance of the nine variables but also their unique and error variances. To have analyzed only their common variance (as estimated by iterative procedures) would have excluded from the resulting factors the specific elements that give each skill its peculiar individuality and set it apart to some degree from others. In this procedure, the most interesting result of the analysis would have been the percentage of the variance of each test left unanalyzed (denoted h^2) minus the chance variance (denoted r_{iI}). The result of the subtraction would, for each test, have represented the percentage of its variance that was not mere chance and that was not measured by the other eight tests. This is its unique nonchance variance.

As indicated, Davis allowed the factors he obtained to be determined by the common variance, the unique nonchance variance, and the chance variance of the nine skill tests. He then determined the stability of the order of the factors obtained and the likelihood that the factors would recur in successive samples of 421 examinees drawn at random from the population represented by the college freshmen whose scores were used in the analysis.

The factor loadings (correlations between each test and each factor) and the factor and test variances are shown in Table 2. The stability of the order of these factors is exceedingly high, as shown by the data in Table 3. This means that in similar analyses made of scores of successive samples drawn from the same population, it is highly likely that the same factors would appear in the same order.

Whether these factors represent nonchance variance can be determined by testing the significance of the reliability coefficients of the factor scores. When these coefficients were estimated empirically in a sample of 100 cases drawn at random from the 421 examinees, the data showed that the hypothesis that each reliability coefficient was significantly positive was sustained at the .05 level, or better, for factors, I, II, III, VII, and VIII. The data are shown in Table 4.

The psychological meaning of factors I, II, III, VII, and VIII rests mainly on subjective judgment, although Davis has reported empirical data pertaining to factors I and II.[15] The best evidence indicates that the variables measured by the five significant factors may be described as follows:

Factor I:	Knowledge of word meanings;
Factor II:	Verbal reasoning;
Factor III:	Sensitivity to implications;
Factor VII:	Following the structure of a passage;
Factor VIII:	Recognizing the literary techniques of a writer.

Davis's study has been reported in detail because it constitutes the base for the design of the present study, which is described in section 2 of this report.

In 1948 Harris published an analysis of comprehension of literature.[16] He

---

[15] F. B. Davis, "Two New Measures of Reading Ability," Journal of Educational Psychology, XXXIII (1942), 365-372.

[16] C. W. Harris, "Measurement of Comprehension of Literature," The School Review, LVI (1948), 280-289; 332-342.

TABLE 2

FACTOR LOADINGS (CORRELATIONS BETWEEN EACH TEST AND EACH FACTOR) FOUND BY DAVIS IN 1941*

Skill	Factor I	II	III	IV	V	VI	VII	VIII	IX	Skill Variance
1	.971	-.235	.016	-.007	-.014	-.001	-.002	.000	.001	134.699
2	.785	.182	.005	-.002	.585	-.011	-.008	-.012	-.023	10.563
3	.456	.149	.002	.000	-.000	.000	.877	.000	.002	3.009
4	.339	.208	.000	.000	.119	.000	.000	.000	.905	1.220
5	.603	.289	-.182	-.003	-.017	.717	-.009	-.014	-.005	6.050
6	.833	.395	-.294	-.215	-.044	-.064	-.012	-.000	-.004	32.169
7	.802	.477	.364	.003	-.049	-.006	-.013	-.022	-.005	33.752
8	.582	.270	.002	.174	-.000	.000	-.008	.738	-.004	3.456
9	.794	.297	-.265	.472	-.038	-.056	-.010	.057	-.003	16.540
Variance	192.270	22.824	8.657	5.282	3.828	3.306	2.327	1.956	1.006	

---

*Frederick B. Davis, "Fundamental Factors of Comprehension in Reading," *Psychometrika*, IX (1944), 185-197.

TABLE 3

VARIANCE RATIOS BETWEEN SUCCESSIVE PAIRS OF FACTOR VARIANCES FOUND BY DAVIS IN 1941*

Factor	dof	Variance	F-ratio#
I	406	192.270	
			8.280
II	399	22.824	
			2.663
III	403	8.657	
			1.622
IV	399	5.282	
			1.387
V	401	3.828	
			1.158
VI	401	3.306	
			1.428
VII	403	2.327	
			1.181
VIII	400	1.956	
			1.944
IX	400	1.006	

*Frederick B. Davis, "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (1944), 185-197.

#With dof = 400, 400, $F_{.10}$ = 1.18 and $F_{.02}$ = 1.27.

TABLE 4

RELIABILITY COEFFICIENTS, MEANS, AND STANDARD DEVIATIONS OF FIVE FACTORS FOUND BY DAVIS IN 1941*

Factor	Reliability Coefficient	Mean	Variance	N
I	.94	46.30	192.37	100
II	.48	24.14	22.85	100
III	.28	.81	8.64	100
VII	.33	.27	2.34	100
VIII	.29	.70	1.96	100

---

*Frederick B. Davis, "Fundamental Factors of Comprehension in Reading," *Psychometrika,* IX (1944), 185-197. Table 7.

prepared objective items to measure seven aspects of comprehension. Fourteen different passages were then chosen and several items were written to measure, with respect to each passage, as many as possible of the seven specified aspects of comprehension. Factor analyses indicated that the common variance of comprehension of the 14 passages (taken in two sets of seven each) could be attributed to one general comprehension skill. That is, the passages did not call for use of different combinations of skills for their comprehension.

He then obtained seven skill scores for each individual across the first and across the second set of passages. Again, factor analysis of the common variance of the intercorrelations of skill scores for each set of passages showed that only one general ability to comprehend (the "verbal factor," perhaps) seemed to be brought into play. Resolution of the variance of each skill score into common variance, unique variance, and chance variance indicated no appreciable amounts of unique variance.

A valuable contribution with respect to the validity of comprehension tests that are made up of short passages followed by questions was provided in 1953 by Derrick.[17] He found that three skills of comprehension in reading were measured about equally well whether the passages on which the questions were based were rather long or short. This is an important point to establish empirically because, to increase efficiency of measurement, the tests used in this study (and in most others) tend to be rather short. Yet, to judge from Derrick's results, conclusions drawn from them may be generalized to the reading of long passages or stories.

A different approach to the study of comprehension skills was used by Lyman C. Hunt in a report published in 1957.[18] He made a differential item analysis of

---

[17] C. Derrick, Three Aspects of Reading Comprehension as Measured by Tests of Different Lengths. Ph.D. dissertation, University of Chicago, Chicago, Illinois, 1953.

[18] L. C. Hunt, Jr., "Can We Measure Specific Factors Associated With Reading Comprehension?" Journal of Educational Research, LI (1957), 161-171.

204 multiple-choice items constructed by F. B. Davis and drawn from the Cooperative Reading Comprehension Tests. Twenty-one judges classified the items with respect to six of the nine skills specified by Davis in 1941. Each skill was represented by 36 items.

In a sample of 370 examinees, Hunt obtained the point-biserial coefficients between individual item score on each of the 204 items and each of the six skill scores. The item-total coefficient for each of the 36 items with the total score of which it was a part was corrected to remove its self-correlation. These corrected coefficients were then adjusted to remove the effect of the unreliability of total score. The resulting item-total coefficients may be regarded as point-biserial coefficients corrected for self-correlation (not merely for spurious correlation of errors) and then corrected for attenuation. The average of the 36 "twice-corrected" coefficients with each skill score was obtained. These averages were then compared. Only the vocabulary items correlated (on the average) significantly more closely with the total score representing their skill (vocabulary) than with total scores representing the other five skills.

Hunt, therefore, concluded that only the vocabulary items were measuring a skill in comprehension (knowledge of word meanings) that was significantly different from the others. This implies that comprehension in reading involves two skills: "word knowledge" and "paragraph comprehension." These results are in harmony with Davis's finding that "word knowledge" and "reasoning in reading" account for virtually all of the variance of comprehension. It is not surprising that tiny components of unique variance in the items should be lost in the approximation procedures used for item analysis, for correction for self-correlation in item-total coefficients, and for correction for attenuation

Alshan in 1964 adopted a slightly different approach to analyzing the skills involved in comprehension.[19] He computed product-moment correlation coefficients of item scores among items 1-40, a complete scale, in Form 2A of the *Davis Reading Test*.[20] This test was constructed to measure five reading skills, as follows:

---

[19] L. M. Alshan, A Factor-Analytic Study of Items Used in the Measurement of Some Fundamental Factors of Reading Comprehension. Ed.D. Report, Teachers College, Columbia University, 1964.

1. Finding answers to questions answered explicitly or in paraphrase in a passage;
2. Weaving together the ideas of a passage and grasping its central thought;
3. Making inferences about the content of a passage and about the purpose or point of view of its author;
4. Recognizing the tone, mood, and literary devices used in a passage;
5. Following the structure of a passage.

The first 40 items in Form 2A included the following item types: Skill 1, 8; Skill 2, 12; Skill 3, 13; Skill 4, 3; Skill 5, 4.

Alshan performed a principal-axis factor analysis of the matrix of item intercorrelations (phi coefficients) and rotated the largest five of the 40 factors by the normalized varimax procedure. The factor loadings of the items of each of Skills 1-5 did not substantiate the hypothesis that five independent mental abilities were being measured by these 40 items.

There were several reasons why this result may have been obtained: First, the individual item reliability coefficients must necessarily have been very small; Second, phi coefficients reflect in their magnitudes the difficulty levels of the items correlated as well as the magnitude of their reliability coefficients and of their underlying true relationships; Third, the number of items for each skill was small (from 3-12); Fourth, the skills measured by each item are overlapping in such a way that the skill it is intended to measure is not systematically preponderant over other skills. Alshan considered all of these matters in his report.

This review of the more important empirical studies of comprehension in

---

[20] F. B. Davis and C. C. Davis, *The Davis Reading Tests, Series 2*. New York: Psychological Corporation, 1962.

reading over the last half century shows that, to be useful, the studies must use carefully written and selected items to measure differential skills and that large samples of examinees and highly refined statistical techniques must be employed if tiny amounts of variance unique to each skill in a set are to be successfully detected.

In the writer's judgment, none of the statistical techniques that has been employed has been sufficiently refined to be capable of detecting small packets of unique variance. For this reason, he designed the study reported in the next section of this report to use for the first time a cross-validated uniqueness analysis based on large samples of examinees. This technique, applied to carefully written and administered variables freed for the first time from spurious elements that have characterized the measurement of reading comprehension, has provided what is intended to be a definitive study of comprehension skills among mature readers.

Section 2

IDENTIFICATION AND MEASUREMENT OF READING SKILLS OF HIGH-SCHOOL STUDENTS

Design of the Study and Procedures Followed

Purpose.

The primary purpose of the study was to obtain estimates of the percentage of nonchance unique variance in the reliable variance of each of the most important measureable skills of comprehension among mature readers.

The Sample.

It was decided to sample the population of twelfth-grade pupils in academic high schools. For preliminary tryout of the test items in March 1966, approximately 400 seniors were tested in Collingdale and Penncrest High Schools in suburban Philadelphia.[21]

---

[21] The writer is grateful to Dr. Harry Heiges and to Dr. Stanley Campbell, the superintendents of schools in the two cooperating school districts, and to the principals and faculty members of the schools.

For the final testing in November 1966, on the results of which the uniqueness analyses were made, approximately 1,100 seniors were tested in Collingdale, Penncrest, Radnor, and Upper Darby High Schools. The pupils tested in Collingdale were juniors when the tryout tests were administered to seniors in that school. There was, therefore, no overlapping of pupils in the tryout and final testing sessions.[22]

---

[22] The writer is grateful to Dr. Harry Heiges, Superintendent of Schools in Collingdale; to Dr. Stanley Campbell, Superintendent of Schools in Rose Tree; to Dr. T. Edward Rutter, Superintendent of Schools in Radnor; and to Dr. H. Curwen Schlosser, Superintendent of Schools in Upper Darby. The cooperation of the principals of the four high schools and of their faculty members is also acknowledged with sincere thanks.

The Comprehension Skills To Be Measured.

After consideration of the experimental studies on the analysis of comprehension in reading, with special attention to the results of studies made since Davis's 1941 study, eight skills were selected for measurement. These are listed in Table 6; a sample item measuring each of the skills is shown in Table 5.

Comparison of the lists of nine skills in Tables 1 and 6 shows that the only important difference between the skills measured by Davis in 1940-1941 and those measured in the present study lies in the combination of Skills 4 and 6 in the 1941 list into Skill 3 in the present study.

Tryout Tests, Forms A and B.

It was decided to try out 40 items of each of the eight skills decided upon. To eliminate the spurious interrelationships among item scores for items based on the same passage, it was determined that each of the 320 items to be tried out should be based on a separate passage. As everyone who has written reading items knows, difficult though it is to write high-quality items measuring the desired skills of comprehension, it is even more difficult and time-consuming to find interesting and meaty passages about which to write the items.

To provide the required 320 items for this study, permission was obtained from the Educational Testing Service for the writer to use items in experimental forms that he prepared in constructing the Cooperative Reading Comprehension Tests, Lower and Higher Levels, Forms Q, R, S, T, Y, and Z. These provided an item pool of nearly 2,000 items, many of which were never used in the published forms of the tests. Similarly, permission was obtained from the Psychological Corporation to use items in experimental forms that C. C. Davis and the writer prepared for the Davis Reading Tests, Series 1 and 2, Forms A, B, C, and D. These provided an additional item pool of about 1,000 items, many of which were not used in the

TABLE 5

SAMPLE ITEMS MEASURING EIGHT SKILLS OF COMPREHENSION IN READING*

1. Remembering Word Meanings

1. guffaw

A make fun of
B sneeze
C cough
D laugh

2. Inferring Word Meanings From Context

Into the muddy pool of my heart some healing drops had fallen--from the music of the passing birds, from the crimson disc that had now dropped below the horizon, the darkening hills, the rose and blue of infinite heaven; and I felt purified and had a strange apprehension of a secret innocence and spirituality in nature--a foreknowledge of some bourn, incalculably distant perhaps, to which we are all moving.

2. "apprehension" (line 8)
A fear
B perception
C recollection
D seizure

3. Understanding Content Stated Explicitly

All program changes must be recorded on blanks furnished by the Registrar and filed with him after they have been approved by the student's adviser, or, in the case of applicants for advanced degrees, by the Director of the School of Education or the Dean of the College of Liberal Arts.

3. Program changes are to be filed with the
A Registrar.
B student's adviser.
C Director of the School of Education.
D Dean of the College of Liberal Arts.

TABLE 5 (Continued)

4. Weaving Ideas in the Content

One early April I visited a man who had an outdoor swimming pool. The first night my host asked, "Are you a morning plunger?"

Thinking he referred to a tub plunge in a warm bathroom, I glowed and said, "You bet!"

"I'll call for you at seven, then, and we'll go out to the pool."

It was evidently his morning custom, and I wasn't going to have it said that a middle-aged man could outdo me. My visit lasted five days, and I later learned from one to whom my host confided that they were the worst five days he had ever gone through. "But I couldn't be outdone by a mere stripling," he said, "and the boy surely enjoyed it."

4. The writer and his host both

   A liked to swim.
   B disliked swimming.
   C were amused by the other's behavior.
   D misunderstood the other's real feelings.

5. Making Inferences About the Content

The delight Tad had felt during his long hours in the glen faded as he drew near the cabin. The sun was nearly gone and Tad's father was at the woodpile. He was wearing the broadcloth suit that he wore to church and to town sometimes. Tad saw his father's hands close around a bundle of wood. He was doing Tad's work--and in his good clothes. Tad ran to him. "I'll git it, Pa."

5. When Tad saw his father, he felt

   A disappointed.
   B impatient.
   C angry.
   D guilty.

TABLE 5 (Continued)

6. Recognizing the Author's Tone, Mood, and Purpose

The golf links lie so near the mill
    That almost every day
The laboring children can look out
    And see the men at play.

6. This verse was written about 1915 and refers to a social problem of the period--child labor. The tone of the verse is

A resigned.
B belligerent.
C bitterly ironic.
D mournful.

7. Identifying the Author's Literary Techniques

Thomas Girard once remarked of George V: "King George does not reign; he merely sprinkles."

7. Girard was making use of

A exaggeration.
B understatement.
C a play on the word "reign."
D a play on the word "sprinkles."

8. Following the Structure of the Content

Only the adult male cricket chirps. On a summer night, they sing by the thousand in unison, so that the forest seems to pulsate and the tiny unseen orchestra becomes its very voice.

8. A "adult male cricket."
B "summer night."
C "forest."
D "tiny unseen orchestra."

---

*Some of these test materials were drawn by permission from the Cooperative Reading Comprehension Tests and the Davis Reading Tests by permission of the Educational Testing Service and the Psychological Corporation, the copyright holders. Their use was authorized only for experimental use in this particular research project.

published forms of the tests.[23]

It should be emphasized at this point that the results of this study depend basically on the appropriateness of the items used. No statistical manipulation of data resulting from the use of items lacking intrinsic validity can wholly make up for their fundamental inadequacy. This fact cannot be overemphasized; basically, this study stands or falls on the psychological insight and ingenuity that characterizes the items used. The outline of Forms A and B is shown in Table 6.

__Experimental Tryout of Forms A and B__.

In March 1966, two parallel forms (A and B) were administered to approximately 400 twelfth-grade pupils on successive mornings. Abundant time was provided for even very slow readers to try every item, but a few did not do so. Careful observation of the pupils led to the identification of a few who were judged not to be trying hard. Their answer sheets and those of pupils who did not try (not necessarily mark) every item were not used for item-analysis purposes.

The distributions of total scores for Forms A and B for the 351 pupils used for item analysis are shown in Table 7. It is interesting to note that the estimate of the reliability of each test (by Kuder-Richardson equation 20) was .96 for each form. This indicates that the reliability coefficient of the total reading-comprehension scores (the sum of scores on Forms A and B) that were reported to the cooperating schools was even higher, making the scores by far the most accurate measures of comprehension in reading ever used.

__Differential Item Analysis__.

The biserial correlation coefficients between pass or fail on each item in

---

[22] The permissions applied only to experimental use of the items for research purposes in this study. The items may not be used again and all copies of the experimental tests have been destroyed. Sample items are provided in Table 5 of this report.

TABLE 6

COMPOSITION OF EXPERIMENTAL READING TESTS, FORMS A AND B, ADMINISTERED IN MARCH 1966

Part	Reading Skill	Number of Items	
		Form A	Form B
1	Recalling word meanings	20	20
2	Finding answers to questions answered explicitly or merely in paraphrase in the content	20	20
3	Weaving together ideas in the content	20	20
4	Drawing inferences from the content	20	20
5	Recognizing a writer's purpose, attitude, tone, and mood	20	20
6	Drawing inferences about the meaning of a word from context	20	20
7	Identifying a writer's techniques	20	20
8	Following the structure of a passage	20	20
	TOTAL	160	160

TABLE 7

DISTRIBUTIONS OF TOTAL SCORES ON FORMS A AND B*

Raw Score	Number of Pupils Form A	 Form B
30-39	4	6
40-49	5	12
50-59	16	17
60-69	22	26
70-79	36	26
80-89	32	42
90-99	43	51
100-109	34	41
110-119	48	48
120-129	44	46
130-139	45	25
140-149	16	11
150-159	6	0
Range	33-154#	31-147#
Mean	102.21	97.40
s	27.36	26.29
KR #20 Homogeneity Coefficient	.96	.96
s_{meas}	5.24	5.31
N	351	351

---

*These data were kindly supplied by Dr. Leonard M. Alshan, City College, City University of New York.

#The most likely chance score is 40.

Form A and total scores for each of the eight skills in Form B were computed. Likewise, the biserial correlation coefficients between pass or fail on each item in Form B and total scores for each of the eight skills in Form A were computed. Thus, each of the 320 items was correlated with eight skill scores in such a way that no coefficient was spuriously increased by inclusion of an item in the total score with which it was correlated.[24] The percentage of the 351 examinees who marked each item correctly was also obtained.

From the 40 items measuring each skill, 24 were selected that had a higher average correlation with the total score on that skill than with the total scores on the other seven skills. The results of this process have been summarized in Table 8. In general, the mean differences are small, as would be expected.[25] Nonetheless, the 24 items for each skill chosen for use in the uniqueness analysis had both subjective and empirical justification for their inclusion.

The median difficulty indexes for these groups of items (expressed as percentages of examinees marking the items correctly) are presented in Table 9. The medians varied from 45 for the vocabulary items to 80 for the items testing ability to draw inferences from the content. These differences in median difficulty indexes were not sufficiently great to be troublesome. The mean scores on Forms C and D were 65.168 and 65.458 in the entire sample used for intercorrelation purposes.

---

[24] The nature of spurious correlation was explicitly dealt with by the writer in 1958. See: F. B. Davis, "A Note on Part-Whole Correlation," *Journal of Educational Psychology*, XLIX (1958), 77-79.

[25] As a matter of fact, they tend to be systematically a little too small because they are averages of correlation coefficients instead of transformations of the coefficients into Fisher's z values. This transformation was not made largely because the coefficients were low (where the transformation makes little difference) and the extra labor could not be justified by any practical outcome.

TABLE 8

MEDIAN CORRELATION COEFFICIENTS OF ITEMS WITH TOTAL SCORES FOR SKILLS SUPPOSED TO BE MEASURED AND FOR SKILLS NOT SUPPOSED TO BE MEASURED*

Skill	Median Correlation Coefficient of 24 Items With Total Score in Skill Supposed To Be Measured By Items	Median Correlation Coefficient of 168 Items With Total Scores in Skills Not Supposed To Be Measured By Items
1	.33	.29
2	.34	.29
3	.37	.31
4	.34	.31
5	.32	.28
6	.36	.32
7	.38	.36
8	.42	.36

*Item-total correlation coefficients are all cross-validated; that is, each item in Form A was correlated with skill total scores in Form B, and *vice versa*. Thus, no coefficients are spuriously high, but all are directly comparable.

TABLE 9

MEDIAN PERCENTAGE OF EXAMINEES MARKING ITEMS CORRECTLY IN FORMS C AND D

Skill	Median Percentage*
1	45
2	67
3	76
4	70
5	80
6.	58
7	60
8	62

*The median percentage of examinees that would be most likely to mark items correctly by chance alone is 25.

These means and the accompanying variances for Forms C and D are shown in Table 12.

Final Tests, Forms C and D.

After a pool of 24 items for each of eight skills had been obtained, the items of each type were assigned to Forms C and D to provide an essentially equivalent pair of tests to measure each of the eight skills. The salient characteristics of these sixteen tests may be summarized as follows:

1. The items in each skill test were initially judged to measure the skill for which they were to be used;
2. All items in each skill test showed a higher biserial correlation coefficient with the skill they were intended to measure than with other skills;
3. All items were of appropriate difficulty for use with twelfth-grade pupils;
4. Each item was based on a separate passage and no passage occurred more than once.

The last point is especially important and makes Tests C and D unique in the measurement of comprehension skills. All previous experimental studies have failed to provide unequivocably separate measures of each skill tested.

The outline for Tests C and D is shown in Table 10.

Administration of Tests C and D.

Tests C and D were administered to approximately 1,100 twelfth-grade pupils. All pupils were scheduled to take both forms and were allowed time enough to try (not necessarily mark) each item. The testing sessions were given with one to two days intervening. Some pupils were absent from one of the testing sessions and some did not try all items. There were also some pupils whose scores differed so widely from day to day that the differences could not reasonably be ascribed to chance. Some of them were observed to be inattentive at one or both of the sessions; others showed no overt signs of failure to cooperate, but their obtained

TABLE 10

COMPOSITION OF FINAL READING TESTS, FORMS C AND D, ADMINISTERED IN OCTOBER-NOVEMBER 1966

Part	Reading Skill	Number of Items Form A	 Form B
1	Recalling word meanings	12	12
2	Drawing inferences about the meaning of a word from context	12	12
3	Finding answers to questions answered explicitly or merely in paraphrase in the content	12	12
4	Weaving together ideas in the content	12	12
5	Drawing inferences from the content	12	12
6	Recognizing a writer's purpose, attitude, tone, and mood	12	12
7	Identifying a writer's techniques	12	12
8	Following the structure of a passage	12	12
	TOTAL	96	96

scores on Tests C and D differed by 12 points or more. Differences of this magnitude, or greater, would occur by chance fewer than five times out of every one hundred. Consequently, answer sheets of this type were excluded from the analyses of data.[26]

---

[26] Special reports to the cooperating schools were made concerning the comprehension skills of these pupils. The standard error of measurement of the difference between obtained scores was found to be 5.5 points from data given in the last line of Table 12.

After removal of all incomplete and unsatisfactory answer sheets, data for 988 pupils were available for analysis. With the answer sheets arranged alphabetically within school, every other pair of answer sheets was assigned to Sample 1 (494 examinees) and Sample 2 (494 examinees). Sample 3 consists of answer sheets from all 988 examinees.

Data From Administration of Tests C and D.

For each examinee, the following scores were obtained:

Form C, score on odd-numbered items in each of Skills 1-8;
Form C, score on even-numbered items in each of Skills 1-8;
Form C, score on all items in each of Skills 1-8;
Form C, score on all items;
Form D, score on odd-numbered items in each of Skills 1-8;
Form D, score on even-numbered items in each of Skills 1-8;
Form D, score on all items in each of Skills 1-8;
Form D, score on all items;
Forms C and D, score on all items.

Thus, 51 scores were available for each examinee. The total scores on Form C, Form D, and Forms C and D combined were reported to the cooperating schools. Except for the last named, the scores were intercorrelated to form a 50-by-50 matrix of product-moment correlation coefficients, bordered by means and variances, for sample 1, for sample 2, and for sample 3.

Reliability Coefficients of Scores on Tests C and D.

The data from the three 50-by-50 matrixes make possible the estimation of reliability coefficients by several methods. Two basic types have been chosen for use in the data analyses reported here. The first type consists of within-day, within-test coefficients to accompany within-test intercorrelations. Table 11 presents these estimates of reliability for Test C and for Test D in each of the three samples. These coefficients were obtained by the procedure originally described by Angoff by means of an equation provided by Davis.[27] This procedure in large measure overcomes the well-known tendency for the Spearman-Brown formula to ~~under~~overestimate the reliability coefficient of a test if the variances of its component half scores are not essentially identical in the sample used. In sample 3 (988 cases), the least reliable skill scores were those of Skill 2 (Drawing inferences about word meanings from context) and the most reliable were those of Skill 8 (Following the structure of a passage).

It should be noted that the within-day reliability coefficients of the sixteen half scores in Test C and the sixteen half scores in Test D are embedded in the three 50-by-50 matrixes and can be used if desired.

Reliability estimates for the eight skill scores in each of Tests C and D can also be obtained across day; that is, from testing session to testing session. These estimates are appropriate for accompanying across-day intercorrelations of Tests C and D. Table 12 presents these across-day reliability estimates separately

---

[27] W. H. Angoff, "Test Reliability and Effective Test Length," Psychometrika, XVIII (1953), 1-14.

F. B. Davis, Educational Measurements and Their Interpretation. Belmont, Calif.: Wadsworth Publishing Co., 1964. Equation F.5, p. 344.

TABLE 11

WITHIN-DAY RELIABILITY COEFFICIENTS*, MEANS, AND VARIANCES OF SKILLS 1-8 IN FORMS C AND D

Sample 1 (N = 494)							
Test	r_{nN}	$\bar{X}_o$	v_o	$\bar{X}_e$	v_e	$\bar{X}_t$	v_t
Form C							
1	.679	3.399	1.822	2.933	2.265	6.332	6.173
2	.497	4.559	1.614	3.623	2.271	8.182	5.143
3	.674	4.700	1.488	4.419	1.781	9.114	4.920
4	.615	3.806	1.949	4.368	1.742	8.774	5.328
5	.682	5.079	1.403	4.763	1.260	9.842	4.039
6	.659	3.634	2.212	3.188	1.994	6.822	6.268
7	.715	3.441	2.206	3.935	2.482	7.376	7.289
8	.766	4.370	2.120	4.079	2.324	8.449	7.201
Form D							
1	.625	3.065	1.898	2.759	1.777	5.824	5.346
2	.609	4.382	1.879	4.105	1.554	8.488	4.927
3	.723	4.964	1.706	4.328	1.876	9.291	5.606
4	.690	4.358	1.986	4.229	2.006	8.587	6.096
5	.671	4.824	1.256	4.714	1.559	9.538	4.224
6	.671	3.698	2.344	3.415	2.028	7.113	6.571
7	.687	3.716	2.381	4.089	2.166	7.806	6.923
8	.648	3.777	1.577	4.040	2.383	7.818	5.804

TABLE 11 (Continued)

Sample 2 (N = 494)

Test	r_{nN}	$\bar{X}_o$	v_o	$\bar{X}_e$	v_e	$\bar{X}_t$	v_t
				Form C			
1	.700	3.490	1.714	2.964	2.160	6.453	5.940
2	.614	4.619	1.534	3.769	2.129	8.389	5.256
3	.602	4.866	1.199	4.601	1.587	9.468	3.969
4	.651	3.872	1.673	4.453	1.712	8.326	5.019
5	.635	5.168	1.077	4.866	1.215	10.034	3.355
6	.654	3.690	2.165	3.269	1.974	6.960	6.148
7	.742	3.632	2.452	4.158	2.181	7.789	7.347
8	.742	4.468	2.103	4.152	2.084	8.619	6.654
				Form D			
1	.571	3.340	2.022	2.929	1.745	6.269	5.268
2	.680	4.579	1.607	4.154	1.631	8.733	4.906
3	.698	5.061	1.550	4.395	1.497	9.455	4.682
4	.746	4.415	2.068	4.465	1.754	8.880	6.085
5	.668	4.976	1.143	4.858	1.395	9.834	3.802
6	.669	3.931	2.189	3.488	1.970	7.419	6.247
7	.664	3.804	2.373	4.136	1.983	7.939	6.511
8	.706	3.751	1.631	4.170	2.401	7.921	6.182

TABLE 11 (Continued)

Sample 3 (N = 988)

Test	r_{nN}	$\bar{X}_o$	v_o	$\bar{X}_e$	v_e	$\bar{X}_t$	v_t
				Form C			
1	.689	3.444	1.769	2.948	2.211	6.393	6.054
2	.557	4.589	1.574	3.696	2.204	8.285	5.205
3	.644	4.783	1.349	4.510	1.691	9.294	4.471
4	.632	3.839	1.811	4.411	1.728	8.250	5.174
5	.660	5.123	1.241	4.815	1.239	9.938	3.703
6	.656	3.662	2.188	3.229	1.984	6.891	6.207
7	.728	3.536	2.336	4.046	2.342	7.583	7.354
8	.754	4.419	2.112	4.115	2.203	8.534	6.928
				Form D			
1	.602	3.202	1.977	2.844	1.767	6.046	5.351
2	.644	4.481	1.751	4.130	1.592	8.610	4.927
3	.713	5.012	1.629	4.361	1.686	9.373	5.146
4	.717	4.387	2.027	4.347	1.892	8.734	6.106
5	.671	4.900	1.205	4.786	1.481	9.686	4.031
6	.670	3.815	2.279	3.451	1.999	7.266	6.426
7	.676	3.760	2.377	4.112	2.073	7.872	6.715
8	.677	3.764	1.603	4.105	2.394	7.869	5.990

*Estimated by Davis equation F.5. Cf. F. B. Davis, Educational Measurements and Their Interpretation. Belmont, Calif.: Wadsworth Publishing Co., 1964, p. 344.

TABLE 12

ACROSS-DAY RELIABILITY COEFFICIENTS*, MEANS, AND VARIANCES OF SKILLS 1-8 IN FORM C OR FORM D

Test	Reliability Coefficient	$\bar{X}_C$	v_C	$\bar{X}_D$	v_D
Sample 1 (N = 494)					
1	.580	6.332	6.173	5.824	5.346
2	.617	8.182	5.143	8.488	4.927
3	.607	9.119	4.920	9.291	5.606
4	.663	8.174	5.328	8.587	6.096
5	.636	9.842	4.039	9.538	4.224
6	.664	6.822	6.268	7.113	6.571
7	.665	7.376	7.289	7.806	6.923
8	.679	8.449	7.201	7.818	5.804
1-8	.935	64.298	232.867	64.466	234.752
Sample 2 (N = 494)					
1	.585	6.453	5.940	6.269	5.268
2	.658	8.389	5.256	8.733	4.906
3	.638	9.468	3.969	9.455	4.682
4	.624	8.326	5.019	8.880	6.085
5	.542	10.034	3.355	9.834	3.802
6	.600	6.960	6.148	7.419	6.247
7	.666	7.789	7.347	7.939	6.511
8	.676	8.619	6.654	7.921	6.182
1-8	.930	66.038	214.118	66.451	219.266

TABLE 12 (continued)

Sample 3 (N = 988)					
1	.582	6.393	6.054	6.046	5.351
2	.639	8.285	5.205	8.610	4.927
3	.621	9.294	4.471	9.373	5.146
4	.644	8.250	5.174	8.734	6.106
5	.594	9.938	3.703	9.686	4.031
6	.633	6.891	6.207	7.266	6.426
7	.665	7.583	7.354	7.872	6.715
8	.677	8.534	6.928	7.869	5.990
1-8	.933	65.168	224.025	65.458	227.766

*Product-moment correlation coefficients between scores on Form C and Form D with interval of 1-3 days between administrations.

for each of the three samples along with the means and variances of the skill scores. As noted in the table, the reliability estimates are raw product-moment coefficients; they have not been adjusted by means of the Angoff equations since the differences between the variances of the skill scores in Tests C and D are too small to make the adjustments of practical consequence.

It is interesting to note that the least reliable of the skill scores were those for Skill 1 (Memory for word meanings) and the most reliable were those for Skill 8 (Following the structure of a passage). As expected, the across-day reliability coefficients of the skill scores tend to be lower than the within-day coefficients.

The across-day reliability coefficient for the total scores on Test C (96 items) and for Test D (96 items) is .93. This cannot be directly compared with the estimated within-day reliability of .96 for the 160 items in Test A or in Test B obtained from data yielded by Kuder-Richardson equation 20.

The close equivalence of total scores from Tests C and D is shown in Table 12 by their almost identical variances and the difference of only .3 point between their means (where the range of possible scores is 0-96).

Intercorrelations of Skill Scores.

As mentioned previously, the three 50-by-50 matrixes of intercorrelations can be broken down into many subsections. However, for purposes of the uniqueness analyses to be made at this time, four basic sets of intercorrelations have been chosen for use. These are:

1. The intercorrelations of the eight skill scores in Test C in each of samples 1, 2, and 3;
2. The intercorrelations of the eight skill scores in Test D in each of samples 1, 2, and 3;
3. The intercorrelations of the eight skill scores in Test C with the eight skill scores in Test D, Test C coming first (that is, $r_{(C1)(D2)}$; $r_{(C1)(D3)}$; . . . ; $r_{(C7)(D8)}$);

4. The intercorrelations of the eight skill scores in Test C with the eight skill scores in Test D, Test D coming first (that is, $r_{(D1)(C2)}$; $r_{(D1)(C3)}$; . . . . ; $r_{(D7)(C8)}$.

These 12 matrixes are shown in Table 13. The values are shown to five decimal places because they were used in multiple-regression analyses.

Multiple-Regression Analyses.

The first step in making a uniqueness analysis is to estimate the variance of each one of a set of variables that can be predicted from (and therefore overlaps) an optimally weighted combination of all other variables in the set. The statistic that represents the desired variance is the squared multiple correlation coefficient between each variable in a set and a combination of all other variables in the set. To obtain this squared coefficient, the proper weight (denoted β) for each predictor variable is determined and used in equation 1 shown in Table 14.

In this study, equation 1 was used eight times for each of the 12 matrixes of intercorrelations shown in Table 13 and summarized for convenience in Table 15. The 12 sets of regression weights were stored for use in cross validation; the squared multiple-regression coefficients in the original samples used for computing the weights are shown in column 4 of Table 16. The first eight of these in Table 16 show, for example, the multiple correlations in matrix 1 (as defined in Table 15) between each of the eight skill scores and the remaining seven skill scores. To illustrate, the squared multiple correlation of Skill 1 (Memory for word meanings) and the optimally weighted combination of the other seven skills is .4383. Consequently, the proportion of the variance of Skill 1 that overlaps the variances of all the other seven skills is .4383 (about 44 per cent) in this sample of 494 examinees.

It can easily be shown, however, that this sample estimate of overlap of variance is spuriously high as an estimate of the overlap of true variance because the multiple-regression procedure uses error components in the sample as well as

TABLE 13

INTERCORRELATIONS OF SKILLS 1-8 IN FORM C*
(WITHIN FIRST TESTING SESSION)

Skill	1	2	3	4	5	6	7	8
				Sample 1 (N = 494)				
1	1.00000	.55656	.46238	.53488	.42642	.51916	.60721	.54311
2	.55656	1.00000	.57062	.60766	.52295	.61946	.64831	.62878
3	.46238	.57062	1.00000	.58694	.56926	.51989	.61189	.60737
4	.53488	.60766	.58694	1.00000	.52442	.59886	.66151	.63208
5	.42642	.52295	.56926	.52442	1.00000	.49824	.53612	.57765
6	.51916	.61946	.51989	.59886	.49824	1.00000	.67397	.60035
7	.60721	.64831	.61189	.66151	.53612	.67397	1.00000	.68012
8	.54311	.62878	.60737	.63208	.57765	.60035	.68012	1.00000
				Sample 2 (N = 494)				
1	1.00000	.59894	.50513	.53530	.39493	.53268	.57606	.52727
2	.59894	1.00000	.57204	.62095	.52999	.60327	.64119	.61224
3	.50513	.57204	1.00000	.59743	.51243	.53923	.61021	.56316
4	.53530	.62095	.59743	1.00000	.49693	.58074	.63059	.56728
5	.39493	.52999	.51243	.49693	1.00000	.43569	.49044	.55908
6	.53268	.60327	.53923	.58074	.43569	1.00000	.64758	.57348
7	.57606	.64119	.61021	.63059	.49044	.64758	1.00000	.64125
8	.52727	.61224	.56316	.56728	.55908	.57348	.64125	1.00000
				Sample 3 (N = 988)				
1	1.00000	.57794	.48218	.53544	.41188	.52614	.59170	.53575
2	.57794	1.00000	.57151	.61465	.52669	.61172	.64564	.62093
3	.48218	.57151	1.00000	.59179	.54526	.52865	.61259	.58691
4	.53544	.61465	.59179	1.00000	.51201	.59032	.64646	.60123
5	.41188	.52669	.54526	.51201	1.00000	.46878	.51537	.56939
6	.52614	.61172	.52865	.59032	.46878	1.00000	.66074	.58754
7	.59170	.64564	.61259	.64646	.51537	.66074	1.00000	.66107
8	.53575	.62093	.58691	.60123	.56939	.58754	.66107	1.00000

*Means and variances of the variables in these samples are given in Table 11. .

TABLE 13

INTERCORRELATIONS OF SKILLS 1-8 IN FORM D*
(WITHIN SECOND TESTING SESSION)

Skill	1	2	3	4	5	6	7	8
Sample 1 (N = 494)								
1	1.00000	.46214	.49178	.52549	.47283	.53586	.50279	.46613
2	.46214	1.00000	.60345	.64816	.64204	.59088	.59966	.58137
3	.49178	.60345	1.00000	.66350	.66537	.62514	.64331	.65077
4	.52549	.64816	.66350	1.00000	.66818	.61948	.65448	.63687
5	.47283	.64204	.66537	.66818	1.00000	.63399	.63146	.60723
6	.53586	.59088	.62514	.61948	.63399	1.00000	.63809	.60701
7	.50279	.59966	.64331	.65448	.63146	.63809	1.00000	.63913
8	.46613	.58137	.65077	.63687	.60723	.60701	.63913	1.00000
Sample 2 (N = 494)								
1	1.00000	.48378	.43552	.50760	.38709	.47138	.45959	.44232
2	.48378	1.00000	.64711	.69281	.59603	.61194	.62407	.60349
3	.43552	.64711	1.00000	.68811	.58041	.60215	.57770	.63325
4	.50760	.69281	.68811	1.00000	.64698	.62132	.63977	.66086
5	.38709	.59603	.58041	.64698	1.00000	.56781	.55076	.56335
6	.47138	.61194	.60215	.62132	.56781	1.00000	.58913	.61792
7	.45959	.62407	.57770	.63977	.55076	.58913	1.00000	.64662
8	.44232	.60349	.63325	.66086	.56335	.61792	.64662	1.00000
Sample 3 (N = 988)								
1	1.00000	.47531	.46564	.51899	.43506	.50668	.48166	.45382
2	.47531	1.00000	.62431	.67152	.62091	.60254	.61189	.59264
3	.46564	.62431	1.00000	.67510	.62636	.61465	.61254	.64128
4	.51899	.67152	.67510	1.00000	.65908	.62170	.64735	.64893
5	.43506	.62091	.62636	.65908	1.00000	.60384	.59285	.58484
6	.50668	.60254	.61465	.62170	.60384	1.00000	.61452	.61225
7	.48166	.61189	.61254	.64735	.59285	.61452	1.00000	.64275
8	.45382	.59264	.64128	.64893	.58484	.61225	.64275	1.00000

*Means and variances of the variables in these samples are given in Table 11.

TABLE 13

INTERCORRELATIONS OF SKILLS 1-8 IN FORMS C AND D*
(FIRST vs SECOND TESTING SESSION;
C1 D2; C1 D3; . . .; C2 D3; C2 D4; . . .; C7 D8)

Skill	1	2	3	4	5	6	7	8
Sample 1 (N = 494)								
1	1.00000	.53507	.50757	.54940	.51025	.55744	.51405	.47400
2	.53507	1.00000	.61977	.62091	.62032	.61016	.61167	.59821
3	.50757	.61977	1.00000	.62376	.60246	.61186	.58328	.58742
4	.54940	.62091	.62376	1.00000	.60992	.63114	.62238	.61151
5	.51025	.62032	.60246	.60992	1.00000	.49833	.55145	.55994
6	.55744	.61016	.61186	.63114	.49833	1.00000	.63208	.58140
7	.51405	.61167	.58328	.62238	.55145	.63208	1.00000	.63733
8	.47400	.59821	.58742	.61151	.55994	.58140	.63733	1.00000
Sample 2 (N = 494)								
1	1.00000	.56994	.48882	.62846	.53916	.55908	.57454	.54381
2	.56994	1.00000	.56648	.64627	.57847	.60580	.59657	.62345
3	.48882	.56648	1.00000	.64779	.58339	.59962	.56176	.57046
4	.62846	.64627	.64779	1.00000	.57192	.58878	.56725	.54681
5	.53916	.57845	.58339	.57192	1.00000	.46330	.50553	.53721
6	.55908	.60580	.59962	.58878	.46330	1.00000	.59139	.58838
7	.57454	.59657	.56176	.56725	.50553	.59139	1.00000	.66868
8	.54381	.62345	.57046	.54681	.53721	.58838	.66868	1.00000
Sample 3 (N = 988)								
1	1.00000	.55265	.49881	.58875	.52435	.55854	.54382	.50921
2	.55265	1.00000	.59407	.63457	.60061	.60890	.60441	.61126
3	.49881	.59407	1.00000	.63585	.59591	.60745	.57283	.57752
4	.58875	.63457	.63585	1.00000	.59219	.61111	.59598	.57911
5	.52435	.60061	.59591	.59219	1.00000	.48311	.53010	.54811
6	.55854	.60890	.60745	.61111	.48311	1.00000	.61239	.58503
7	.54382	.60441	.57283	.59598	.53010	.61239	1.00000	.65278
8	.50921	.61126	.57752	.57911	.54811	.58503	.65278	1.00000

---

*Means and variances of the variables in these samples are given in Table 12.

TABLE 13

INTERCORRELATIONS OF SKILLS 1-8 IN FORMS C AND D*
(FIRST vs SECOND TESTING SESSION;
D1 C2; D1 C3; . . .; D2 C3; D2 C4; . . .; D7 C8)

Skill	1	2	3	4	5	6	7	8
				Sample 1 (N = 494)				
1	1.00000	.50242	.45454	.47585	.42566	.49632	.56332	.54564
2	.50242	1.00000	.55740	.59218	.53646	.58172	.66272	.59986
3	.45454	.55740	1.00000	.63901	.55906	.58702	.67763	.68707
4	.47585	.59218	.63901	1.00000	.59213	.63117	.65682	.70214
5	.42566	.53646	.55906	.59213	1.00000	.62805	.64542	.64593
6	.49632	.58172	.58702	.63117	.62805	1.00000	.75022	.63007
7	.56332	.66272	.67763	.65682	.64542	.75022	1.00000	.70268
8	.54564	.59986	.68707	.70214	.64593	.63007	.70268	1.00000
				Sample 2 (N = 494)				
1	1.00000	.52087	.41996	.47203	.33838	.45241	.49557	.47093
2	.52087	1.00000	.61394	.58575	.53466	.59225	.65246	.61372
3	.41996	.61394	1.00000	.61532	.58192	.55385	.65649	.63869
4	.47203	.58575	.61532	1.00000	.56379	.61004	.67910	.66895
5	.33838	.53466	.58192	.56379	1.00000	.53560	.62240	.59878
6	.45241	.59225	.55385	.61004	.53560	1.00000	.65192	.60110
7	.49557	.65246	.65649	.67910	.62240	.65192	1.00000	.59306
8	.47093	.61372	.63869	.66895	.59878	.60110	.59306	1.00000
				Sample 3 (N = 988)				
1	1.00000	.51310	.44212	.47469	.38640	.47483	.53284	.50957
2	.51310	1.00000	.58500	.58958	.53633	.58736	.65887	.60708
3	.44212	.58500	1.00000	.62804	.57018	.57121	.66717	.66462
4	.47469	.58958	.62804	1.00000	.57924	.62096	.66938	.68610
5	.38640	.53633	.57018	.57924	1.00000	.58336	.63599	.62379
6	.47483	.58736	.57121	.62096	.58336	1.00000	.70280	.61654
7	.53284	.65887	.66717	.66938	.63599	.70280	1.00000	.65007
8	.50957	.60708	.66462	.68610	.62379	.61654	.65007	1.00000

---

*Means and variances of the variables in these samples are given in Table 12.

TABLE 14

EQUATIONS USED IN ESTIMATING UNIQUE NONCHANCE VARIANCE

Let: $a, b, \ldots, g$ represent variables in a set administered to a sample drawn at random from a population;

i represent a variable other than $a, b, \ldots, g$ in the same set administered to the same sample;

$a', b', \ldots, g'$ represent the same variables as $a, b, \ldots, g$ administered to a different sample drawn from the same population;

i' represent a variable other than $a', b', \ldots, g'$ administered to the same sample as $a', b', \ldots, g'$;

$\beta_a, \beta_b, \ldots, \beta_g$ represent multiple-regression coefficients in standard-measure form for maximizing the correlation of a combination of variables $a, b, \ldots, g$ with variable i;

$\beta'_a, \beta'_b, \ldots, \beta'_g$ represent multiple-regression coefficients in standard-measure form for maximizing the correlation of a combination of variables $a', b', \ldots, g'$ with variable i';

k represent the weighted combination of variables $a, b, \ldots, g$ that has the highest possible correlation with variable i;

k' represent the weighted combination of variables $a', b', \ldots, g'$ that has the highest possible correlation with variable i';

r^2_{ik} represent the proportion of the variance of variable i that overlaps the variance of composite k (This is the squared multiple correlation coefficient of variables $a, b, \ldots, g$ with variable i);

$r^2_{ik'}$ represent the proportion of the variance of variable i that overlaps the variance of composite k' (This is the squared cross-validated multiple correlation coefficient of variables $a', b', \ldots, g'$ with variable i);

r_{iI} represent the reliability coefficient of variable i;

$r_{k'K'}$ represent the reliability coefficient of composite k';

s^2_{Ui} represent the unique nonchance variance of variable i in the set of variables administered to the sample to which variable i was administered.

${}_cs^2_{Ui'}$ represent the cross-validated proportion of unique variance in the nonerror variance of variable i.

TABLE 14 (Continued)

$$r_{ik}^2 = \sum^{g} \beta_a r_{ai} \tag{1}$$

$$r_{ik'}^2 = \frac{\left(\sum_{a'}^{g'} \beta'_a \, r_{ia} \right)^2}{\sum_{a'}^{g'} {\beta'_a}^2 + C} \tag{2}$$

$$r_{k'K'} = \frac{\sum_{a'}^{g'} {B'_a}^2 \, r_{ai} + C}{\sum_{a'}^{g'} {\beta'_a}^2 + C} \tag{3}$$

$${}_c s_{Ui'}^2 = \frac{r_{iI} \, r_{k'K'} - r_{ik'}^2}{r_{iI} \, r_{k'K'}} \tag{4}$$

where:

$$C = \sum_{a=1}^{g'} \sum_{b=1}^{g'} \beta'_a \, \beta'_b \, r_{ab} \qquad [a \neq b]$$

TABLE 15

MATRIXES OF INTERCORRELATIONS USED FOR ESTIMATING MULTIPLE CORRELATIONS

Matrix Number	Sample	Test Form	Description
1	1	C	Within first testing session
2	1	D	Within second testing session
3	1	C, D	First vs second testing session
4	1	D, C	Second vs first testing session
6	2	C	Within first testing session
7	2	D	Within second testing session
8	2	C, D	First vs second testing session
9	2	D, C	Second vs first testing session
11	3	C	Within first testing session
12	3	D	Within second testing session
13	3	C, D	First vs second testing session
14	3	D, C	Second vs first testing session

TABLE 16

PROPORTION OF VARIANCE OF EACH OF EIGHT VARIABLES ACCOUNTED FOR BY SEVEN OTHERS IN ORIGINAL AND CROSS-VALIDATED SAMPLES

		Original Samples		Cross-Validated Samples					
Matrix No.*	Criterion	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2
1	1	1	.4383	2	.4099	6	.4282	7	.4202
1	2	1	.5616	2	.5350	6	.5566	7	.5465
1	3	1	.5113	2	.5051	6	.5004	7	.4888
1	4	1	.5597	2	.5498	6	.5544	7	.5446
1	5	1	.4395	2	.4227	6	.4288	7	.4098
1	6	1	.5431	2	.5174	6	.5376	7	.5166
1	7	1	.6545	2	.6426	6	.6525	7	.6370
1	8	1	.5977	2	.5890	6	.5921	7	.5854
2	1	2	.3693	1	.3439	6	.3469	7	.3611
2	2	2	.5422	1	.5204	6	.5221	7	.5339
2	3	2	.6039	1	.5953	6	.5968	7	.5878
2	4	2	.6222	1	.6130	6	.6154	7	.6207
2	5	2	.6022	1	.5797	6	.5610	7	.5949
2	6	2	.5690	1	.5440	6	.5496	7	.5639
2	7	2	.5883	1	.5793	6	.5831	7	.5778
2	8	2	.5596	1	.5529	6	.5435	7	.5569
3	1	3	.4233	4	.3594	8	.4026	9	.3658
3	2	3	.5752	4	.5479	8	.5589	9	.5554
3	3	3	.5545	4	.4990	8	.5303	9	.5249
3	4	3	.5895	4	.5686	8	.5480	9	.5608
3	5	3	.5216	4	.4394	8	.5142	9	.4447
3	6	3	.5669	4	.5063	8	.5602	9	.5383
3	7	3	.5659	4	.5478	8	.5529	9	.5152
3	8	3	.5367	4	.5125	8	.5205	9	.4834

TABLE 16 (Continued)

		Original Samples		Cross-Validated Samples					
Matrix No.*	Criterion	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2
4	1	4	.3804	3	.3410	8	.3366	9	.3667
4	2	4	.5119	3	.4934	8	.4891	9	.5004
4	3	4	.5699	3	.5316	8	.5192	9	.5582
4	4	4	.5955	3	.5785	8	.5296	9	.5786
4	5	4	.5262	3	.4485	8	.4296	9	.5103
4	6	4	.6213	3	.5528	8	.5331	9	.6000
4	7	4	.7156	3	.6964	8	.6741	9	.6857
4	8	4	.6653	3	.6388	8	.5998	9	.6451
6	1	6	.4554	1	.4444	2	.4271	7	.4404
6	2	6	.5856	1	.5800	2	.5599	7	.5673
6	3	6	.5040	1	.4915	2	.4964	7	.4876
6	4	6	.5416	1	.5360	2	.5319	7	.5295
6	5	6	.4045	1	.3935	2	.3740	7	.3677
6	6	6	.5207	1	.5146	2	.4963	7	.4962
6	7	6	.6126	1	.6110	2	.6041	7	.5963
6	8	6	.5494	1	.5439	2	.5265	7	.5229
7	1	7	.3227	1	.3082	2	.3161	6	.3146
7	2	7	.5978	1	.5819	2	.5865	6	.5804
7	3	7	.5794	1	.5546	2	.5638	6	.5595
7	4	7	.6628	1	.6461	2	.6610	6	.6521
7	5	7	.5016	1	.4749	2	.4961	6	.4675
7	6	7	.5372	1	.5104	2	.5323	6	.5160
7	7	7	.5494	1	.5340	2	.5402	6	.5381
7	8	7	.5784	1	.5678	2	.5755	6	.5580
8	1	8	.5143	3	.4909	4	.4356	9	.4394
8	2	8	.5788	3	.5642	4	.5492	9	.5449
8	3	8	.5565	3	.5354	4	.4880	9	.4995
8	4	8	.6019	3	.5645	4	.5281	9	.5001
8	5	8	.4809	3	.4756	4	.3926	9	.3958
8	6	8	.5342	3	.5285	4	.4617	9	.5024
8	7	8	.5598	3	.5474	4	.5243	9	.4655
8	8	8	.5673	3	.5498	4	.5006	9	.4560

TABLE 16 (Continued)

		Original Samples		Cross-Validated Samples					
Matrix No.*	Criterion	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2
9	1	9	.3453	3	.3001	4	.3336	8	.2930
9	2	9	.5617	3	.5414	4	.5448	8	.5213
9	3	9	.5637	3	.5413	4	.5514	8	.5123
9	4	9	.5964	3	.5784	4	.5750	8	.5123
9	5	9	.4953	3	.4297	4	.4759	8	.4083
9	6	9	.5301	3	.5015	4	.5057	8	.4915
9	7	9	.6500	3	.6023	4	.6174	8	.5616
9	8	9	.5988	3	.5538	4	.5763	8	.5158
11	1	11	.4440	12	.4258				
11	2	11	.5723	12	.5561				
11	3	11	.5065	12	.5000				
11	4	11	.5497	12	.5397				
11	5	11	.4219	12	.3976				
11	6	11	.5305	12	.5074				
11	7	11	.6337	12	.6242				
11	8	11	.5729	12	.5574				
12	1	12	.3464	11	.3328				
12	2	12	.5675	11	.5527				
12	3	12	.5880	11	.5801				
12	4	12	.6422	11	.6334				
12	5	12	.5524	11	.5253				
12	6	12	.5538	11	.5330				
12	7	12	.5668	11	.5594				
12	8	12	.5670	11	.5551				
13	1	13	.4609	14	.4030				
13	2	13	.5741	14	.5556				
13	3	13	.5510	14	.5054				
13	4	13	.5863	14	.5460				
13	5	13	.5014	14	.4239				
13	6	13	.5502	14	.5088				
13	7	13	.5599	14	.5195				
13	8	13	.5466	14	.4913				

TABLE 16 (Continued)

		Original Samples		Cross-Validated Samples					
Matrix No.*	Criterion	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2	Betas From Matrix No.*	Multiple r^2
14	1	14	.3613	13	.3248				
14	2	14	.5327	13	.5157				
14	3	14	.5640	13	.5314				
14	4	14	.5917	13	.5630				
14	5	14	.5078	13	.4356				
14	6	14	.5704	13	.5231				
14	7	14	.6771	13	.6398				
14	8	14	.6278	13	.5806				

*See Table 15 for description of matrix used.

true-score components. This overestimation of overlap can most satisfactorily be eliminated by using a set of regression coefficients determined on one set of data with the intercorrelation matrix obtained by using the same tests in a different sample of examinees or by using equivalent (but different) tests in the same sample of examinees. Both techniques have been employed in this study. These techniques are commonly referred to as "techniques of cross validation."

The basic equation for obtaining the squared multiple correlation between a given variable and an optimally weighted set of other variables in a different sample of examinees is equation 2 in Table 14. Columns 6, 8, and 10 of Table 16 show the squared multiple correlations of each skill score with the optimally weighted and cross-validated combination of the seven other skill scores in the set. The matrix of intercorrelations used in obtaining the data on each line of Table 16 is shown in column 1. The matrixes used to compute the regression weights for the squared multiple correlations in columns 6, 8, and 10 are shown in columns 5, 7, and 9, respectively.

As noted previously, the matrixes are identified for convenience in Table 15. A glance at the data on each line in Table 16 shows the amount of overestimation in the sample estimates of overlapping variance that appear in column 4. For the first entry in the table, the overlapping variance of Skill 1 and the other seven variables is reduced from .4383 to .4099 if matrix 2 is used for cross validation (that is, if different but equivalent items are given to the same examinees); from .4383 to .4282 if matrix 6 is used for cross validation (that is, if the same items are given at the same time to a different random sample of examinees drawn from the same population); and from .4383 to .4202 if matrix 7 is used for cross validation (that is, if different but equivalent items are given to a different random sample of examinees drawn from the same population).

This same type of comparison may be made for the data on each of the remaining 95 lines in Table 16. Inspection of the data will show that, for samples 1 and 2, cross validation has been accomplished separately by exchanging examinees, by exchanging items, and by exchanging both examinees and items simultaneously. For sample 3, cross validation has been limited to exchanging items only. It may be worth pointing out at this time that the design of this study makes it possible to cross validate by items alone, by examinees alone, or by a combination of these in sample 3 if half scores are used in place of entire skill scores. Only the added complexity of presenting the data has led to a decision to confine this presentation to studies of entire skill scores.

Uniqueness Analysis.

It is clear that the maximum proportion of obtained variance that one variable can have in common with other variables of defined types can be determined by making use of the cross-validated multiple-regression procedures previously described.

Let: s^2_{0i} represent the total obtained variance of variable i;

r^2_{ik} represent the squared multiple correlation of variable 1 and weighted composite k (where the weights maximize the value of r^2_{ik});

s^2_{Ui} represent the unique variance of variable i (that is, the part of the total variance of variable i that does not overlap the total variance of composite k).

Then: $s^2_{0i} = r^2_{ik} + s^2_{Ui}$.

In standard measures, $s^2_{0i} = 1$. Hence,

$$1 = r^2_{ik} + s^2_{Ui};$$

or: $s^2_{Ui} = 1 - r^2_{ik}$.

If we add prime marks to k and Ui to show that they represent cross-validated data, we obtain:

$$s^2_{Ui'} = 1 - r^2_{ik'} .$$

Equation 2 in Table 14 may be used to estimate $r^2_{ik'}$.

To remove the influence of unreliability in variable i and composite k' from $r^2_{ik'}$, this coefficient can be corrected for attenuation by dividing it by the product of the reliability coefficients of variable i (denoted r_{iI}) and of composite k' (denoted $r_{k'K'}$). The appropriateness of the reliability coefficients so used is crucial. Each one must appropriately accompany the intercorrelations in the matrix used to compute $r^2_{ik'}$ and $r_{k'K'}$. The time interval between parallel tests used in computing the reliability coefficients should be the same as the time interval between administration of the tests in the matrix. The reliability coefficient of composite k' may be obtained in each instance by using equation 3 in Table 14.

After correction for attenuation, we have equation 4 in Table 14:

$$ {}_c s^2_{Ui'} = 1 - \frac{r^2_{ik'}}{r_{iI}\, r_{k'K'}} \; ; \text{ or} $$

$$ {}_c s^2_{Ui'} = \frac{r_{iI}\, r_{k'K'} - r^2_{ik'}}{r_{iI}\, r_{k'K'}} $$

A formal derivation of this equation has been presented by Horst.[28] A different form of this equation along with others was given earlier by Shaycoft.[29]

It should be clear that ${}_c s^2_{Ui'}$ is a cross-validated estimate of the proportion of the reliable variance of variable i that is independent of the variance of any element in composite k'. This is defined by Horst as an estimate of specificity and is appropriate for general use in comparing the proportions of unique nonchance variance in measures of unequal reliability.

---

[28] A. P. Horst, Psychological Measurement and Prediction. Belmont, California: Wadsworth Publishing Co., 1966. Equation (20.5.25), p. 333.

[29] M. F. Shaycoft, "What the Tests Measure," in J. C. Flanagan, F. B. Davis, et al., The American High-School Student. Pittsburgh: Project TALENT Office, University of Pittsburgh, 1964.

If one wants to estimate the proportion of total obtained variance of variable i that is unique nonchance variance so far as the elements of composite k' are concerned, a different estimate should be made. This requires the correction of $r^2_{ik'}$ for the unreliability of composite k' alone. If we denote this estimate of unique variance by a double prime mark, we may write:

$$ {}_{c}s^2_{Ui''} = 1 - \frac{r^2_{ik'}}{r_{k'K'}} ; \quad \text{or} $$

$$ {}_{c}s^2_{Ui''} = \frac{r_{k'K'} - r^2_{ik'}}{r_{k'K'}} . $$

This cross-validated estimate of unique variance is of practical use in the selection of tests to make up an efficient battery for selection or classification purposes. As far as the writer is aware, it was first so used by Flanagan.[30]

---

[30] J. C. Flanagan, Flanagan Aptitude Classification Tests: Technical Report. Chicago: Science Research Associates, 1959.

For the uniqueness estimates in this study, equation 4 in Table 14 was employed. As noted, this is identical with Horst's estimate of specificity. The squared multiple correlations used are shown in columns 6, 8, and 10 of Table 16 and the reliability coefficients used are given in Table 17. The latter were obtained by equation 3 in Table 14. Because they represent combinations of 84 items, they tend to take rather high values (close to .90 in this situation).

Results of the Study

Estimates of Uniqueness.

Table 18 presents the complete set of 224 estimates of the proportion of unique variance in the nonerror variance of each of the eight skills in comprehension. For each skill, these estimates are separated into two groups: those

TABLE 17

RELIABILITY COEFFICIENTS OF READING-SKILL SCORES AND OF CROSS-VALIDATED WEIGHTED COMPOSITE SCORES (OF SEVEN VARIABLES EACH) USED IN ESTIMATING UNIQUE NONCHANCE VARIANCE

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
1	2	1	.8938	.6793
		2	.9073	.4973
		3	.9208	.6735
		4	.9164	.6154
		5	.9008	.6822
		6	.9235	.6587
		7	.9201	.7147
		8	.9102	.7663
1	6	1	.8788	.6793
		2	.9233	.4973
		3	.9161	.6735
		4	.9115	.6154
		5	.8755	.6822
		6	.9044	.6587
		7	.9125	.7147
		8	.9112	.7663
1	7	1	.8764	.6793
		2	.9113	.4973
		3	.8910	.6735
		4	.9112	.6154
		5	.8851	.6822
		6	.9238	.6587
		7	.9081	.7147
		8	.9074	.7663
2	1	1	.8928	.6255
		2	.9239	.6091
		3	.9061	.7227
		4	.9220	.6903
		5	.9049	.6706
		6	.8942	.6709
		7	.9133	.6871
		8	.9246	.6479

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
2	6	1	.8926	.6255
		2	.9211	.6091
		3	.9203	.7227
		4	.9195	.6903
		5	.8717	.6706
		6	.9023	.6709
		7	.9146	.6871
		8	.9166	.6479
2	7	1	.8912	.6255
		2	.9231	.6091
		3	.9033	.7227
		4	.9184	.6903
		5	.9041	.6706
		6	.9224	.6709
		7	.9041	.6871
		8	.9160	.6479
3	4	1	.8672	.5804
		2	.8977	.6172
		3	.8797	.6069
		4	.8973	.6628
		5	.8929	.6363
		6	.8559	.6643
		7	.8977	.6649
		8	.8985	.6787
3	8	1	.8745	.5804
		2	.9059	.6172
		3	.8690	.6069
		4	.8658	.6628
		5	.8563	.6363
		6	.8809	.6643
		7	.8877	.6649
		8	.8823	.6787

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient Composite Score	Skill Score
3	9	1	.8439	.5804
		2	.9031	.6172
		3	.8975	.6069
		4	.8948	.6628
		5	.8733	.6363
		6	.9018	.6643
		7	.8907	.6649
		8	.8827	.6787
4	3	1	.9060	.5804
		2	.9184	.6172
		3	.9142	.6069
		4	.9194	.6628
		5	.8734	.6363
		6	.8887	.6643
		7	.9100	.6649
		8	.9083	.6787
4	8	1	.8801	.5804
		2	.9107	.6172
		3	.8780	.6069
		4	.8596	.6628
		5	.8564	.6363
		6	.8843	.6643
		7	.8904	.6649
		8	.8867	.6787
4	9	1	.8487	.5804
		2	.9080	.6172
		3	.9020	.6069
		4	.9016	.6628
		5	.8816	.6363
		6	.9056	.6643
		7	.8856	.6649
		8	.8765	.6787

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
6	1	1	.9029	.6996
		2	.9232	.6140
		3	.9030	.6019
		4	.9240	.6512
		5	.8926	.6348
		6	.9034	.6542
		7	.9130	.7417
		8	.9181	.7415
6	2	1	.8947	.6996
		2	.9016	.6140
		3	.9167	.6019
		4	.9164	.6512
		5	.9043	.6348
		6	.9217	.6542
		7	.9166	.7417
		8	.9071	.7415
6	7	1	.8873	.6996
		2	.9088	.6140
		3	.8980	.6019
		4	.9091	.6512
		5	.8923	.6348
		6	.9226	.6542
		7	.9107	.7417
		8	.9065	.7415
7	1	1	.8979	.5714
		2	.9217	.6800
		3	.9107	.6985
		4	.9209	.7458
		5	.9103	.6678
		6	.8977	.6693
		7	.9158	.6643
		8	.9231	.7065

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
7	2	1	.9082	.5714
		2	.9194	.6800
		3	.9228	.6985
		4	.9204	.7458
		5	.9228	.6678
		6	.9174	.6693
		7	.9283	.6643
		8	.9195	.7065
7	6	1	.9061	.5714
		2	.9170	.6800
		3	.9225	.6985
		4	.9178	.7458
		5	.8901	.6678
		6	.9054	.6693
		7	.9198	.6643
		8	.9123	.7065
8	3	1	.8881	.5851
		2	.9037	.6583
		3	.8996	.6378
		4	.9089	.6240
		5	.8766	.5422
		6	.8874	.6000
		7	.9014	.6664
		8	.8992	.6758
8	4	1	.8693	.5851
		2	.8944	.6583
		3	.8746	.6378
		4	.8945	.6240
		5	.8841	.5422
		6	.8471	.6000
		7	.8884	.6664
		8	.8923	.6758

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
8	9	1	.8476	.5851
		2	.9026	.6583
		3	.8937	.6378
		4	.8933	.6240
		5	.8706	.5422
		6	.9004	.6000
		7	.8810	.6664
		8	.8739	.6758
9	3	1	.8909	.5851
		2	.9045	.6583
		3	.9004	.6378
		4	.9088	.6240
		5	.8760	.5422
		6	.8869	.6000
		7	.9014	.6664
		8	.9037	.6758
9	4	1	.8685	.5851
		2	.8948	.6583
		3	.8787	.6378
		4	.8963	.6240
		5	.8878	.5422
		6	.8533	.6000
		7	.8860	.6664
		8	.8904	.6758
9	8	1	.8624	.5851
		2	.8949	.6583
		3	.8548	.6378
		4	.8643	.6240
		5	.8619	.5422
		6	.8864	.6000
		7	.8852	.6664
		8	.8832	.6758

TABLE 17 (Continued)

Correlation Matrix No.	Beta Weight Matrix No.	Skill Estimated	Reliability Coefficient	
			Composite Score	Skill Score
11	12	1	.8936	.6889
		2	.9111	.5568
		3	.9146	.6439
		4	.9141	.6325
		5	.9004	.6605
		6	.9248	.6565
		7	.9188	.7276
		8	.9090	.7543
12	11	1	.9026	.6016
		2	.9229	.6435
		3	.9177	.7127
		4	.9222	.7167
		5	.8993	.6706
		6	.9017	.6701
		7	.9168	.6759
		8	.9211	.6773
13	14	1	.8667	.5822
		2	.9078	.6387
		3	.8913	.6213
		4	.9041	.6441
		5	.8876	.5938
		6	.8871	.6329
		7	.9003	.6654
		8	.8958	.6770
14	13	1	.8969	.5822
		2	.9143	.6387
		3	.8955	.6213
		4	.9047	.6441
		5	.8723	.5938
		6	.8903	.6329
		7	.9017	.6654
		8	.9005	.6770

TABLE 18

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 1

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.2827	.3249	.2941
494	.2965	.3841	.3991
494	.3522	.2906	.3786
494	.3908	.3923	.3178
Average	.3306	.3480	.3474
988		.3083	
988		.3871	
Average		.3477	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.2067	.2861	.2532
494	.0554	.3517	.4243
494	.2554	.1139	.3410
494	.3436	.4191	.1435
Average	.2153	.2927	.2905
988		.2011	
988		.3780	
Average		.2896	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 2

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	-.2123	.1859	-.2059
494	.0503	.0752	.0693
494	-.0233	-.0168	-.0116
494	.0619	.0691	.0716
Average	-.0309	-.0146	-.0192
988		-.0962	
988		-.0694	
Average		-.0134	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0004	.0110	.0036
494	.1299	.1297	.1071
494	.0516	.0830	.0673
494	.0752	.1151	.0907
Average	.0643	.0847	.0672
988		.0417	
988		.1170	
Average		.0794	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 3

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.1890	.1856	.1855
494	.0996	.0910	.0693
494	.0959	.0977	.1004
494	.1253	.1318	.1283
Average	.1275	.1265	.1209
988		.1510	
988		.1131	
Average		.1321	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	-.0055	.0654	.0364
494	-.0197	.0420	.0255
494	.0669	.1237	.1251
494	.0162	.0602	.0575
Average	.0145	.0728	.0611
988		.0874	
988		.0449	
Average		.0662	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 4

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0118	.0252	.0289
494	.0208	.0369	.0306
494	.1092	.1056	.1086
494	.0370	.0472	.0593
Average	.0447	.0537	.0569
988		.0666	
988		.0418	
Average		.0542	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0450	.0481	.0543
494	.0318	.0505	.0706
494	.0048	.1028	.0539
494	-.0281	.0503	-.0199
Average	.0134	.0629	.0397
988		.0623	
988		.0338	
Average		.0481	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 5

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.2869	.3121	.3215
494	.0188	.0448	.0402
494	.3055	.3510	.3484
494	.1951	.2135	.2186
Average	.2016	.2304	.2322
988		.3314	
988		.1290	
Average		.2302	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0563	.2267	.1997
494	.0866	.1931	.2116
494	-.0006	.1614	.1811
494	.0112	.1263	.0952
Average	.0384	.1769	.1719
988		.1958	
988		.1591	
Average		.1775	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 6

Within-Day Correlation Matrix

		Cross Validation By	
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0975	.1494	.1512
494	.0887	.0932	.0922
494	.1293	.1779	.1768
494	.1331	.1485	.1505
Average	.1122	.1423	.1427
988		.1642	
988		.1180	
Average		.1411	

Across-Day Correlation Matrix

		Cross Validation By	
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0427	.1096	.1015
494	.0027	.0637	.0926
494	.0073	.0700	.0915
494	.0123	.0758	.0577
Average	.0163	.0798	.0858
988		.0939	
988		.0717	
Average		.0828	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 7

Within-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	-.0005	.0228	.0185
494	.0699	.0768	.0721
494	.0978	.1172	.1115
494	.1240	.1195	.1223
Average	.0728	.0841	.0811
988		.0663	
988		.0971	
Average		.0817	

Across-Day Correlation Matrix

	Cross Validation By		
N	Examinees	Day and Items	Examinees, Day, and Items
494	.0634	.0823	.1300
494	-.1387	-.1509	-.1644
494	.0886	.2071	.1144
494	-.0457	.0481	-.0027
Average	-.0081	.0467	.0193
988		.1327	
988		-.0663	
Average		.0332	

TABLE 18 (Continued)

PROPORTIONS OF UNIQUE VARIANCE IN THE NONERROR VARIANCE
SKILL 8

Within-Day Correlation Matrix

N	Cross Validation By Examinees	Day and Items	Examinees, Day, and Items
494	.1519	.1556	.1581
494	.0617	.0771	.0847
494	.2011	.2220	.2174
494	.1141	.1342	.1294
Average	.1322	.1472	.1474
988		.1870	
988		.1103	
Average		.1487	

Across-Day Correlation Matrix

N	Cross Validation By Examinees	Day and Items	Examinees, Day, and Items
494	.0953	.2279	.1698
494	.1308	.1596	.1929
494	-.0846	-.0362	.0033
494	.0422	.1359	.0932
Average	.0459	.1218	.1148
988		.1898	
988		.0476	
Average		.1187	

estimated from within-day correlation matrixes, and those estimated from across-day correlation matrixes. It will be recalled that Form C was administered in the first testing session. The intercorrelations of Skills 1-8 in Form C constitute a matrix of within-day coefficients. Similarly, the intercorrelations of Skills 1-8 in Form D constitute a different matrix of within-day coefficients.

Two across-day correlation matrixes were obtained. The first consists of the correlations of scores on Skill 1 in Form C with the scores on Skill 2 in Form D, on Skill 2 in Form C with Skill 3 in Form D, etc. The second consists of the correlations of scores on Skill 1 in Form D with the scores on Skill 2 in Form C, on Skill 2 in Form D with Skill 3 in Form C, etc.

Cross validation (the use of multiple-regression weights computed in one matrix to obtain multiple correlation coefficients in a different but analogous matrix) yielded the proportions of unique variance shown in three columns in Table 18. The left-hand column in each part of the table shows proportions of unique variance estimated when cross validation was effected by using two different but equivalent sets of items given to the same examinees. The middle column shows proportions estimated when cross validation was effected by using two different but equivalent samples of examinees who were given the same items. The right-hand column shows proportions estimated when cross validation was effected by using two different but equivalent samples of items in two different but equivalent samples of examinees. The data show a tendency for cross validation by examinees along to yield smaller estimates of unique variance than either of the other procedures, one of which (in the right-hand column) included cross validation by examinees.

It seems reasonable to the writer to prefer cross validation by items only. Sampling variation in sets of items drawn from the eight populations of comprehension skills is of chief concern in this study. A summary of uniqueness estimates from within-day and across-day matrixes that result from cross validation by items in the entire sample of 988 examinees is given in Table 19 in terms of percentages (instead of proportions) of unique variance in the nonerror variance of Skills 1-8.

It will be noted that one of these percentages is slightly negative; presumably this is a chance deviation from a true percentage close to zero. The largest percentages of unique variance occur in the case of Skill 1 (Memory for word meanings). The data indicate that about 32 per cent of the nonerror variance of this skill is not involved in any of the other seven comprehension skills used in this study. The second largest percentages of unique variance occurs in the case of Skill 5 (Drawing inferences from the content). About 20 per cent of the nonerror variance of this mental operation is not involved in any of the other seven comprehension skills measured in this study.

It is interesting to compare these results with those obtained in the factor analysis reported by Davis in 1949.[31] These two skills provided the largest factor loadings for the two major components of comprehension, named at that time "Memory for Word Meanings" and "Reasoning in Reading." (See Table 2 in this report.) It is probable that the percentages of unique variance in the nine skills tested in 1940 largely determined the outcome of that analysis since it involved the total variances (not merely the common variances) of the tests.

Three other skills that show appreciable percentages of unique variance are Skill 8 (Following the structure of a passage), Skill 6 (Recognizing a writer's purpose, attitude, tone, and mood), and Skill 3 (Finding answers to questions asked explicitly or in paraphrase).

---

[31] F. B. Davis, "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (1944), 185-197.

TABLE 19

SUMMARY OF PERCENTAGES OF UNIQUE VARIANCE IN THE NONERROR VARIANCE OF EIGHT READING SKILLS

(N = 988)

	CROSS VALIDATION BY	
Skill	Items and Day (Within-Day Matrix)	Items (Across-Day Matrix)
1. Recalling word meanings	35	29
2. Drawing inferences about the meaning of a word from context	-1	8
3. Finding answers to questions answered explicitly or in paraphrase	13	7
4. Weaving together ideas in the content	5	5
5. Drawing inferences from the content	23	18
6. Recognizing a writer's purpose, attitude, tone, and mood	14	8
7. Identifying a writer's techniques	8	3
8. Following the structure of a passage	15	12

Section 3

CONCLUSIONS

Comprehension among mature readers is not a unitary mental skill, or operation. The data summarized in Table 19 leave no doubt that substantial parts of the mental abilities used in eight skills judged to be of importance in comprehension are independent of one another. For example, about 32 per cent of the nonerror variance of a recognition-vocabulary test (Measuring memory for word meanings) was found to be unique in the set of eight skills used in this study. Similarly, about 20 per cent of the nonerror variance of Skill 5 (Drawing inferences about the content of the material read) was found to be unique in the set of eight skills.

The implications of these conclusions for the teaching of reading after the establishment of basic mechanical skills are clear. First, systematic and carefully planned learning exercises that are appropriate in level of difficulty for each pupil should be provided throughout the secondary-school grades. Exercises should be prepared to:

1. Make pupils familiar with the meanings of as many words as possible by means of field trips, visual aids (such as slides, motion pictures, and television), written and oral composition, discussions of books and readings required in various school subjects, and a graded series of passages that introduce the first 30,000 words in frequency in appropriate contexts.
2. Increase the tendency of pupils to draw inferences from what they read and to do this more accurately. The data suggest that weaving ideas and getting the central thought of a passage are subsidiary steps to drawing inferences. In practice, the subsidiary steps should be taught and practiced separately in a series of interesting passages of diverse types and levels of difficulty.
3. Improve pupils' abilities to:
    a. follow the structure of a passage;

b. find answers to questions answered explicitly or in paraphrase in the material read;

c. recognize an author's attitude, tone, mood, and purpose.

This should also be done systematically by providing supervised practice in exercising these skills with interesting passages of diverse types and levels of difficulty.

It is of less consequence whether separate practice exercises are used to improve pupils with respect to Skills 2, 4, and 7. However, practice materials may be used to illustrate these aspects of comprehension, especially Skill 7 (Identifying literary techniques) in the eleventh and twelfth grades.

Unfortunately, learning materials of the types required have not been tried out experimentally and assembled in convenient units. Yet, this study has shown that part of the variance of these skills is unique; therefore, teaching one of them cannot be counted on to cause improvement in others. Clearly, self-teaching exercises of these types should now be prepared and published for secondary-school use.

A comparison of the findings of this study with other empirical studies of comprehension shows broad agreement in most cases. In Davis's study, reported in 1944, the most important factors were interpreted as measures of "Memory for word meanings" and "Reasoning in reading" (a combination of weaving ideas together and drawing inferences from them). Other skills were represented by factors having comparatively small variances. It seems likely that the unique variance associated with each skill largely determined the factor pattern because Davis analyzed the total variance and the common variance of the nine tests he used tended to be largely attributable to one mental trait perhaps best described as verbal reasoning ability.

Although Hunt did not so interpret his results, the writer believes that, within the limits of precision of his data, Hunt found essentially the same two important elements as Davis in comprehension. Harris obtained very little specificity in his tests, but he was concerned mainly with comprehension in literature. His separate tests do not seem to have succeeded in measuring different abilities. Alshan's item intercorrelations did not yield interpretable results, possibly because of the confounding of influences on them, as he pointed out.

None of the studies prior to the present one made use of tests as carefully constructed and as free from spurious effects. Furthermore, none of them used a technique, which like uniqueness analysis, was as sensitive to the presence of small proportions of unique variance. In the writer's judgment, the present study makes it unlikely that additional studies of this type will yield additional information worth heavy expenditures. The next profitable step in the analysis of comprehension skills probably consists of applying these techniques to comprehension in the middle grades (4-9) and in the elementary grades (1-3). New lists of skills will have to be formulated to precede item construction for these studies.

The methodological features of the present study may be helpful in research design. This represents the first large-scale precise application of cross-validated uniqueness-analysis techniques. As such, its capabilities for comparing the effects of cross validating by examinees, by items, or by testing session should be of interest. The matching of reliability coefficients to intercorrelations in terms of length of time interval between testings is worthy of consideration in any study that involves the correction for attenuation. A straightforward test of the significance of the difference between an estimate of unique nonerror variance and zero unique nonerror variance should be used in subsequent studies.

Finally, the most clear-cut finding of this study is the need for a series of self-teaching practice exercises for developing proficiency in the constituent skills of comprehension among mature readers.

BIBLIOGRAPHY

Alderman, G. H., "Improving Comprehension in Reading," Journal of Educational Research, XIII (1926), 11-21.

Alshan, L. M., A Factor-Analytic Study of Items Used in the Measurement of Some Fundamental Factors of Reading Comprehension. Ed.D. report, Teachers College, Columbia University, New York, N. Y., 1964.

Angoff, W. H., "Test Reliability and Effective Test Length," Psychometrika, XVIII (1953), 1-14.

Berry, B. T., "Improving Freshman Reading Habits," English Journal (College Edition), XX (1931), 824-828.

Buswell, G. T., "The Relationship Between Rate of Thinking and Rate of Reading," The School Review, LIX (1951), 339-346.

Carroll, R. P., An Experimental Study of Comprehension in Reading. New York: Bureau of Publications, Teachers College, Columbia University, 1927.

Davis, F. B., Educational Measurements and Their Interpretation. Belmont, Calif.: Wadsworth Publishing Co., 1964.

Davis, F. B., Fundamental Factors of Comprehension in Reading. Ed.D. dissertation, Harvard University, Cambridge, Mass., 1941.

Davis, F. B., "Fundamental Factors of Comprehension in Reading," Psychometrika, IX (1944), 185-197.

Davis, F. B., "A Note on Part-Whole Correlation," Journal of Educational Psychology, XLIX (1958), 77-79.

Davis, F. B., "The New Measures of Reading Ability," Journal of Educational Psychology, XXXIII (1942), 365-372.

Davis, F. B. and Davis, C. C., The Davis Reading Tests, Series 1 and 2, Forms A, B, C, D. New York: Psychological Corporation, 1958; 1962.

Davis, F. B. et al. The Cooperative Reading Comprehension Tests, Lower and Higher Levels, Forms Q, R, S, T, Y, Z. Princeton, N.J.: Educational Testing Service, 1941-1949.

Derrick, C., Three Aspects of Reading Comprehension as Measured by Tests of Different Lengths. Ph.D. dissertation, University of Chicago, Chicago, 1953.

Dewey, J. C., "The Acquisition of Facts as a Measure of Reading Comprehension," Elementary School Journal, XXXV (1935), 346-348.

Feder, D. D., "Comprehension Maturity Tests -- A New Technique in Mental Measurement," Journal of Educational Psychology, XXIX (1938), 597-606.

Flanagan, J. C., Flanagan Aptitude Classification Tests: Technical Report. Chicago: Science Research Associates, 1959.

Gans, R., A Study of Critical Reading Comprehension in the Intermediate Grades. New York: Bureau of Publications, Teachers College, Columbia University, 1940.

Hall, W. E. and Robinson, F. P., "An Analytical Approach to the Study of Reading Skill," Journal of Educational Psychology, XXXVI (1945), 429-442.

Harris, C. W., "Measurement of Comprehension in Literature," The School Review, LVI (1948), 280-289; 332-342.

Horst, A. P. Psychological Measurement and Prediction. Belmont, Calif.: Wadsworth Publishing Co., 1966.

Hunt, L. C., Jr., "Can We Measure Specific Factors Associated With Reading Comprehension?" Journal of Educational Research, LI (1957), 161-171.

Irion, T. W. H., Comprehension Difficulties of Ninth-Grade Students in the Study of Literature. New York: Teachers College, Columbia University, 1925.

Langsam, R. S., "A Factorial Analysis of Reading Ability," Journal of Experimental Education, X (1941), 57-63.

Pickford, R. W., "Some Effects of Style on Reading," British Journal of Psychology, XXVII (1937), 292-296.

Pickford, R. W., "Some Mental Functions Illustrated by an Experiment on Reading," British Journal of Psychology, XXV (1935), 417-434; XXVI (1935), 49-55.

Pickford, R. W., "The Tendency Toward Synthesis in Reading," *British Journal of Psychology*, XXIV (1933), 50-66.

Richards, I. A., *Interpretation in Teaching*. New York: Harcourt, Brace and World, 1938.

Richards, I. A., *Practical Criticism*. New York: Harcourt, Brace and World, 1929.

Shaycoft, M. F., "What the Tests Measure," in Flanagan, J. C., Davis, F. B., *et al.*, *The American High School Student*. Pittsburgh: Project TALENT Office, University of Pittsburgh, 1964.

Thorndike, E. L., "The Psychology of Thinking in the Case of Reading," *Psychological Review*, XXIV (1917), 220-234.

Thorndike, E. L., "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," *Journal of Educational Psychology*, VIII (1917), 323-332.

Thorndike, E. L., "The Understanding of Sentences," *Elementary School Journal*, XVIII (1917), 98-114.

Touton, F. C. and Berry, B. T., "Reading Comprehension at the Junior College Level," *California Quarterly of Secondary Education*, VI (1931), 245-251.

Zahner, L. C., "The Testing of Comprehension," *Educational Record*, Supplement No. 13, XXI (1940), 71-89.